

69206

U.S. DEPARTMENT OF COMMERCE
Patent and Trademark Office

SEARCH REQUEST FORM

Requestor's

Name:

Soubello

Serial

Number:

09/462 955

Date:

6/19

Phone:

308-6043

Art Unit:

1632

Search Topic:

Please write a detailed statement of search topic. Describe specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples or relevant citations, authors, keywords, etc., if known. For sequences, please attach a copy of the sequence. You may include a copy of the broadest and/or most relevant claim(s).

✓ 655-676

682-701

✓ 211-991

✓ 409-991

✓ 611-991

711-991

734-785

of seq ID
NO. 1 B

no pending

STAFF USE ONLY

Date completed:

6/20

Searcher:

S. S. L. 308-6043

Terminal time:

7

Elapsed time:

9

CPU time:

Total time:

Number of Searches:

Number of Databases:

Search Site

STIC

✓ CM-1

Pre-S

Type of Search

7 N.A. Sequence

A.A. Sequence

Structure

Bibliographic

Vendors

IG

STN

Dialog

APS

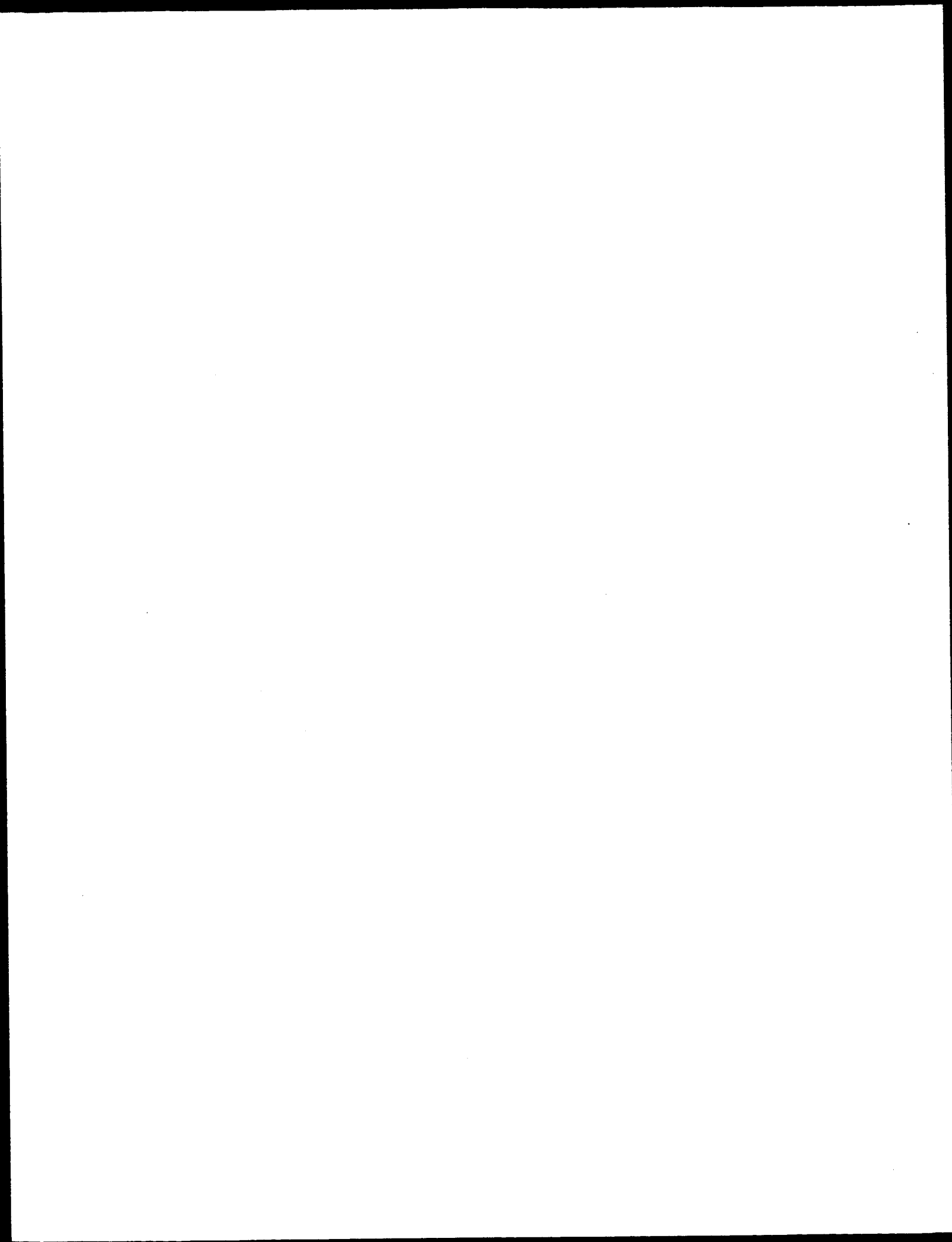
Geninfo

SDC

DARC/Questel

✓ Other

BEST AVAILABLE COPY



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2002, 15:43:14 ; Search time 5904.86 Seconds
(without alignments)
70.879 Million cell updates/sec

Title: US-09-462-955b-1_COPY_682_701
Perfect score: 20
Sequence: 1 ctaccgcggccgaagccggg 20

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl.*

- 1: gb_ba.*
- 2: gb_hgt.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pi.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_hgt_hum.*
- 31: em_hgt_inv.*
- 32: em_hgt_other.*
- 33: em_hgt_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	Score	Match	Length	DB	ID	Description
--------	-------	-------	-------	--------	----	----	-------------

1	20	100.0	21	6	ARI172776
2	20	100.0	1291	6	ARI172770
3	20	100.0	1291	14	CFDCG
4	17.4	87.0	181	8	CISAT
5	17.4	87.0	20400	8	AF271736S1
6	17.4	87.0	21040	8	AF271738S2
7	17.4	87.0	2411	10	AF2717902
8	17	85.0	3830	1	KPN9980
9	17	85.0	31938	2	AC094656
10	17	85.0	64620	2	AC103767
11	16.8	84.0	2142	1	CTV16136
12	16.8	84.0	5451	6	AX078430
13	16.8	84.0	6220	1	AB019552
14	16.8	84.0	6220	1	AB078015
15	16.8	84.0	6220	1	AB078016
16	16.8	84.0	6575	1	AB078018
17	16.8	84.0	6576	1	AB078017
18	16.8	84.0	6576	1	AB078022
19	16.8	84.0	6630	1	AB078019
20	16.8	84.0	6631	1	AB078021
21	16.8	84.0	6632	1	AB019554
22	16.8	84.0	6660	1	AB078020
23	16.8	84.0	10146	1	AE004743
24	16.8	84.0	10717	1	AB008745
25	16.8	84.0	251700	1	AF000062
26	16.4	82.0	990	4	AY011570
27	16.4	82.0	990	4	AY011571
28	16	80.0	341	6	AX184404
29	16	80.0	5135	1	TTNUSA1NF
30	16	80.0	10305	1	AE009820
31	16	80.0	16140	1	AE007014
32	16	80.0	20431	1	MTCY2B12
33	16	80.0	169329	2	AL592211
34	15.8	79.0	337	1	AF012803
35	15.8	79.0	580	8	AF019821
36	15.8	79.0	633	1	AF004392
37	15.8	79.0	979	10	RATPAM28
38	15.8	79.0	1821	8	CEY16012
39	15.8	79.0	2595	1	AB029372
40	15.8	79.0	2614	1	AF227236
41	15.8	79.0	2929	10	RNPAM202
42	15.8	79.0	3212	10	RNPAM203
43	15.8	79.0	3225	6	E03428
44	15.8	79.0	3226	6	AR036184
45	15.8	79.0	3226	10	RNPAM205

ALIGNMENTS

RESULT 1

ARI172776

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

BASE COUNT

ORIGIN

ARI172776 Sequence
ARI172770 Sequence
M25963 Coconut fol
X51946 Citrus icha
AF271736 Euglena g
AF271739 Euglena g
AF277902 Rattus no
AJ009980 Klebsiell
AC094656 Rattus no
AC103767 Homo sapi
Y16136 Clostridium
X51946 Citrus icha
AF271736S1
AF271738S2
AF277902
1 KPN9980
2 AC094656
2 AC103767
1 CTV16136
6 AX078430
1 AB019552
1 AB078015
1 AB078016
1 AB078018
1 AB078017
1 AB078022
1 AB078019
1 AB078021
1 AB019554
1 AB078020
1 AE004743
1 AB008745
1 AF000062
4 AY011570
4 AY011571
6 AX184404
1 TTNUSA1NF
1 AE009820
1 AE007014
1 MTCY2B12
2 AL592211
337 1 AF012803
580 8 AF019821
633 1 AF004392
979 10 RATPAM28
1821 8 CEY16012
2595 1 AB029372
2614 1 AF227236
2929 10 RNPAM202
3212 10 RNPAM203
3225 6 E03428
3226 6 AR036184
3226 10 RNPAM205

21 bp
DNA
linear PAT 17-DEC-2001

Sequence 7 from patent US 6303345.
ARI172776
ARI172776.1 GI:17912267

Unknown.
Unknown.
Unclassified.
1 (bases 1 to 21)
Rohde,W., Becker,D., Randles,J.W., Hehn,A. and Salamini,F.
Use of a virus DNA as promoter
Patent: US 6303345-A 7 16-OCT-2001;
Location/Qualifiers
1. 21
/organism="unknown"

4 a 8 c 8 g 1 t

Query Match 100.0%; Score 20; DB 6; Length 21;
Best Local Similarity 100.0%; Pred. No. 1.5e+03;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctaccggccgaagccggg 20
|||||
Db 1 CTACCCGCCGAAGCCGGG 20

RESULT 2

AR172770 1291 bp DNA linear PAT 17-DEC-2001
LOCUS
DEFINITION Sequence 1 from patent US 6303345.
ACCESSION AR172770
VERSION AR172770.1 GI:17912261
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 1291)
AUTHORS Rohde, W., Becker, D., Randles, J. W., Hehn, A. and Salamini, F.
TITLE Use of a virus DNA as promoter
JOURNAL Patent: US 6303345-A 1 16-OCT-2001;
FEATURES
Location/Qualifiers
1..1291
/organism="unknown"

BASE COUNT 336 a 323 c 332 g 300 t
ORIGIN

Query Match 100.0%; Score 20; DB 6; Length 1291;
Best Local Similarity 100.0%; Pred. No. 3.8e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctaccggccgaagccggg 20
|||||
Db 1072 CTACCCGCCGAAGCCGGG 1091

RESULT 3

CFDCG
LOCUS 1291 bp ss-DNA circular VRL 02-AUG-1993
DEFINITION Coconut foliar decay virus, complete genome.
ACCESSION M29963
VERSION M29963.1 GI:323306
KEYWORDS circular; complete genome.
SOURCE Coconut foliar decay virus
ORGANISM Coconut foliar decay virus

viruses; ssDNA viruses; Nanovirus.
REFERENCE 1 (bases 1 to 1291)
AUTHORS Rohde, W., Randles, J. W., Langridge, P. and Hanold, D.
TITLE Nucleotide sequence of a circular single-stranded DNA associated with coconut foliar decay virus
JOURNAL Virology 176, 648-651 (1990)
MEDLINE 90266484
COMMENT Draft entry and printed sequence for [1] kindly submitted by W. Rhode, 15-MAR-1989, for release after publication.

FEATURES
source
1..1291
/organism="Coconut foliar decay virus"
/db_xref="taxon:12474"
40..70
/note="stem-loop structure"
stem_loop
103..975
CDS
/note="ORF 1"
/codon_start=1
/protein_id="AAA42894.1"
/db_xref="GI:323307"

TRANSLATION="MGSIRWCFNLVETEEAANVVRRESLNLVYIVGDEVAPS
TGQRHQFHLKTRRLOGLKTVLGNDRHLEFTRGSDQNRDYCKSERVLLEHGV
TRPGVKRPLAORFAEEDELRLEDGGYRCVVGASVEMRWAAENPPFFPHNQ
LEVLSAIGEPADRTILWICDGGDKSVFAYLGLKPDWFTCGGTRKDLVLYIE
DPKRLIIDVPRCLNLEYLNALLECVKNRAFSSDKYELPSYLGFDHVLVFNVLDP
YLKISDRDKLWNI"
314..775
CDS

/note="ORF 2"
/codon_start=1
/protein_id="AAA42895.1"
/db_xref="GI:323308"
/translation="MTGFTWSRPVPTNRLETTVRNGCFSTESRLVLESKGDWPN
DLLRLNMSAWKQADETALYELRWNGQDGLPKIRSHFHIIGSLKCCLRSESQRT
IAQSSGYADETEETGSPCLNIDSPTSPTGTHVVEPTTYCTSTSTQNEI"
complement(422..568)
/note="ORF 6"
/codon_start=1
/protein_id="AAA42896.1"
/db_xref="GI:323309"
/translation="MENGTDQRPILSPPLKRVORIGIRLGLPGGVHQPQOIVGP
IVAF"
639..797
/note="ORF 3"
/codon_start=1
/protein_id="AAA42897.1"
/db_xref="GI:323310"
/translation="MTRRRRREVRCVQISRTQARLVLHMMWNOKGRIVPVHRGPKTK
FNPRCTOV"
complement(823..987)
/note="ORF 5"
/codon_start=1
/protein_id="AAA42898.1"
/db_xref="GI:323311"
/translation="MTHTLNIPQFYSVPADFQIIRQDIGKEYMHMVEPKITKGFV
RTECPVLNTF"
1098..1286
/note="ORF 4"
/codon_start=1
/protein_id="AAA42899.1"
/db_xref="GI:323312"
/translation="MNRVMGGPTINDSIWIRTNLLCLQCQPLSTSPQVSSLEKK
AASLVPSICFCAIGRLS"
BASE COUNT 336 a 323 c 332 g 300 t
ORIGIN

Query Match 100.0%; Score 20; DB 14; Length 1291;
Best Local Similarity 100.0%; Pred. No. 3.8e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctaccggccgaagccggg 20
|||||

Db 1072 CTACCCGCCGAAGCCGGG 1091

RESULT 4

CISAT
LOCUS 181 bp DNA linear PLN 29-OCT-1996
DEFINITION Citrus ichangensis satellite DNA.
ACCESSION X51946
VERSION X51946.1 GI:1653975
KEYWORDS extrachromosomal DNA; satellite DNA.
SOURCE Citrus ichangensis.
ORGANISM Citrus ichangensis.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Sapindales; Rutaceae; Citrus.
REFERENCE 1 (bases 1 to 181)
AUTHORS Beridze, T. G.
TITLE Direct Submission
JOURNAL Submitted (21-PEB-1990) Beridze T. G., Institute of Plant
Biochemistry, Georgian Academy of Science, Tbilisi 380031, USSR
REMARK Revised by author 29-OCT-96
REFERENCE 2 (bases 1 to 181)
AUTHORS Beridze, T., Tsirekidze, N. and Roytberg, M. A.
TITLE On the tertiary structure of satellite DNA
JOURNAL Biochimie 74 (2), 187-194 (1992)
MEDLINE 92256540
COMMENT On Oct 30, 1996 this sequence version replaced gi:18037.
FEATURES Location/Qualifiers

source
1. .181
/organism="Citrus ichangensis"
/db_xref="taxon:2709"
/tissue_type="leaves"

misc_feature
40 a 54 c 64 g 23 t
BASE COUNT
ORIGIN

Query Match 87.0%; Score 17.4; DB 8; Length 181;
Best Local Similarity 94.7%; Pred. No. 7.6e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 taccggccggaagccggg 20
Db 26 TACCGCCGCGAAGCCGGG 44

RESULT 5
AF271736S1 20400 bp DNA linear PLN 02-MAY-2001
LOCUS
DEFINITION
Euglena gracilis gamma-tubulin gene, exons 1-14 sequence.
ACCESSION
AF271736
VERSION
AF271736.1 GI:13661562
KEYWORDS
1 of 2
SEGMENT
Euglena gracilis.
SOURCE
Euglena gracilis.
ORGANISM
Eukaryota; Euglenozoa; Euglenida; Euglenales; Euglena.

REFERENCE
1 (bases 1 to 20400)
Canaday,J., Tessier,L.H., Imbault,P. and Paulus,F.
Analysis of Euglena gracilis alpha-, beta- and gamma-tubulin genes:
introns and pre-mRNA maturation
Mol. Genet. Genomics 265 (1), 153-160 (2001)
21263010
JOURNAL
MEDLINE
REFERENCE
2 (bases 1 to 20400)
Paulus,F.P. and Tessier,L.H.
Direct Submission
Submitted (23-MAY-2000) IBMP, CNRS, 12 rue du General Zimmer,
Strasbourg 67084, France
FEATURES
Location/Qualifiers
1. .20400
/organism="Euglena gracilis"
/strain="Klebs Z"
/db_xref="taxon:3039"
814. .869
/number=1
exon
8142. .8254
/number=2
exon
9340. .9393
/number=3
exon
10800. .10913
/number=4
exon
11103. .11171
/number=5
exon
12490. .12569
/number=6
exon
13657. .13702
/number=7
exon
14064. .14150
/number=8
exon
14705. .14791
/number=9
exon
17219. .17368
/number=10
exon
17477. .17548
/number=11
exon
17782. .17856
/number=12
exon
18317. .18384
/number=13
exon
19559. .19652

BASE COUNT 5172 a 4729 c 4709 g 5790 t
ORIGIN

Query Match 87.0%; Score 17.4; DB 8; Length 20400;
Best Local Similarity 94.7%; Pred. No. 1.6e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 taccggccggaagccggg 20
Db 11876 TGCCGCGCGAAGCCGGG 11894

RESULT 6
AF271738S2 21040 bp DNA linear PLN 23-APR-2001
LOCUS
DEFINITION
Euglena gracilis gamma-tubulin gene, exons 2-17, complete cds.
ACCESSION
AF271739
VERSION
AF271739.1 GI:13661567
KEYWORDS
2 of 2
SEGMENT
Euglena gracilis.
SOURCE
Euglena gracilis.
ORGANISM
Eukaryota; Euglenozoa; Euglenida; Euglenales; Euglena.

REFERENCE
1 (bases 1 to 21040)
Canaday,J., Tessier,L.H., Imbault,P. and Paulus,F.
Analysis of Euglena gracilis alpha-, beta- and gamma-tubulin genes:
introns and pre-mRNA maturation
Mol. Genet. Genomics 265 (1), 153-160 (2001)
21263010
JOURNAL
MEDLINE
REFERENCE
2 (bases 1 to 21040)
Paulus,F.P. and Tessier,L.H.
Direct Submission
Submitted (24-MAY-2000) IBMP, CNRS, 12 rue du General Zimmer,
Strasbourg 67084, France
FEATURES
Location/Qualifiers
1. .21040
/organism="Euglena gracilis"
/strain="Klebs Z"
/db_xref="taxon:3039"
join(AF271738.1:1576. .1631,5406. .5518,6546. .6599,
8036. .8149,8339. .8407,9728. .9807,10965. .11010,
11366. .11452,12006. .12092,14512. .14661,14769. .14840,
15074. .15148,15563. .15630,16802. .16895,17682. .17726,
19824. .19894,20826. .21040)
/product="gamma-tubulin"
join(AF271738.1:1583. .1631,5406. .5518,6546. .6599,
8036. .8149,8339. .8407,9728. .9807,10965. .11010,
11366. .11452,12006. .12092,14512. .14661,14769. .14840,
15074. .15148,15563. .15630,16802. .16895,17682. .17726,
19824. .19894,20826. .20937)
/codon_start=1
/product="gamma-tubulin"
/protein_id="AAK37967.1"
/db_xref="GI:13661568"

translation="MPREITLQCGGQGVTFWVKOLCAEHGPNPDGLIODYATILG
GDRKDYEFYQADDEHYIPRALLDLEPRVINTTGEYRNLYNPENIYSEGGGAGN
NWAGYEKADQSDIEDVIDMIDREADGESLEGFLVHSTAGTGGSGSYLLEKLNDR
YPKKLVQTSVFPNDSKESGVVQVYNSLLTKRLTLNADCVVLDNLNRIYVE
RLHPSPFTSQNALSTVMAASTTTLRYPGYNNDLIGLIASLFTTRCHFLMTGYT
PIIDPNIOMTVRKTSVLDVMRLLSPKNIMVSASTRRCYVSILNIQGVDPGOVH
KSLQIRERKMATFIPWGPASIOVALSRKSPYLETARHVSGLMLNHTSIANLFARTL
SNYDKLRKRSALFDVYTKTDLDFKDGLEEDSSREIVONLVEEYKACERPDYINWGSSM
YNAEGGTPAMM"
5406. .5518
/number=2
exon
6546. .6599
/number=3
exon
8036. .8149
/number=4
exon
8339. .8407
/number=5

FEATURES
Location/Qualifiers
1. .21040
/organism="Euglena gracilis"
/strain="Klebs Z"
/db_xref="taxon:3039"
join(AF271738.1:1576. .1631,5406. .5518,6546. .6599,
8036. .8149,8339. .8407,9728. .9807,10965. .11010,
11366. .11452,12006. .12092,14512. .14661,14769. .14840,
15074. .15148,15563. .15630,16802. .16895,17682. .17726,
19824. .19894,20826. .21040)
/product="gamma-tubulin"
join(AF271738.1:1583. .1631,5406. .5518,6546. .6599,
8036. .8149,8339. .8407,9728. .9807,10965. .11010,
11366. .11452,12006. .12092,14512. .14661,14769. .14840,
15074. .15148,15563. .15630,16802. .16895,17682. .17726,
19824. .19894,20826. .20937)
/codon_start=1
/product="gamma-tubulin"
/protein_id="AAK37967.1"
/db_xref="GI:13661568"

```
/db_xref="SPTREMBL:Q9ZEW/"
/translation="GNPMSAVEROOSH LARKKETHKEMRVYVTSEIKDEFRRMCEAQG
```

```

/feature="tapA"
432..509
/feature="tapA"
/function="antisense control"
/codon_start=1
/transl_table=11
/product="Tapa"
/protein_id="CAA08924.1"
/db_xref="GI:4138116"
/db_xref="SPTREMBL:Q9ZEW6"
/translation="MLRKLQAQFLCHWLLLCNLSAVGCD"
502..1365
/feature="repA"
502..1365
/feature="repA"
/function="Replication protein"
/codon_start=1
/transl_table=11
/product="RepA"
/protein_id="CAA08925.1"
/db_xref="GI:4138117"
/db_xref="SPTREMBL:Q9ZEW5"
/translation="MTESLLSHYKQVKNPNVFTPREGKTLTPCRKLMAKAEFT
SRFDSIHVAFSLGKGRHMPPLRRRAIDALLQAMCHYDPLANRVORSTTNMAIE
CRLATSKSNLSITRATRALKELAEGLITYQTEYDPOIGCNIPDTIDFTPALFSAL
DVSEVAVVARRSRVWENKQKQKLPLEMDLAKAWRFVRRFRSVQSRKSHG
RKRTARRDASRQKDIETRVROQLTREVATGRFGDHEALKREVERVQERMLLSRG
NNVRLATVPI"
1535..1719
/feature="oriR"
1535..1719
/direction=RIGHT
1535..1719
/feature="oriR"
1538..1546
/feature="oriR"
/function="DnaA box"
complement(1823..1845)
/feature="TerRleft"
complement(1823..1845)
/feature="TerRleft"
1953..1974
/feature="TerRright"
1953..1974
/feature="TerRright"
2148..2160
/function="proposed site of leading strand DNA synthesis"
complement(2267..2772)
/feature="orf2"
complement(2267..2772)
/feature="orf2"
/codon_start=3
/transl_table=11
/protein_id="CAA08927.1"
/db_xref="GI:4138119"
/db_xref="SPTREMBL:Q9ZEW3"
/translation="ETFISGRDNQKPFRLITLKRASDRCHDWLAVFTVDEIGLL
LVSTLFPKPAISKADPTVLPKRFKHPAAGCLNPGINOWRFIPPPWRTVTPVRIEL
QLFVAFQAQLHPGTRGDIATDGFVRVGNISIGFIEERKILFPRHLRKTSAHGYLQ
WPLKYR"
2558..2974
/feature="orf1"
2558..2974
/feature="orf1"
/feature="homology to yedG gene, D90833"
/codon_start=1
/transl_table=11
/protein_id="CAA08926.1"
/db_xref="GI:4138118"
/db_xref="SPTREMBL:Q9ZEW4"
/translation="WFKPLWQHGRAICFADGWFKWREGDKKQPYFIHRKDGOPILMA
AIGTSPFERGDEAGFLIVTAAADKGLVDIHDRPLVLVLPDARVWVKQDVSGKEAD
IAADGAVSADHFTHWPTVTRVGNVKNQGPFLIEPVG"

```

BASE COUNT

928 a

945 c

999 g

958 t

ORIGIN

Query Match 85.0%; Score 17; DB 1; Length 3830;
 Best Local Similarity 100.0%; Pred. No. 4e+03;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 cccggcggaagcgcg 20
 |||||
 Db 1919 CCCGGCGGAAGCGCGG 1935

RESULT 9
 LOCUS AC094656
 DEFINITION Rattus norvegicus clone CH230-5b20, *** SEQUENCING IN PROGRESS ***,
 20 unordered pieces.
 ACCESSION AC094656
 VERSION 5 GI:17941433
 KEYWORDS HTG; HTGS_PHASE1.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus

REFERENCE 1 (bases 1 to 31938)
 AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
 Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
 Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
 Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
 Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
 Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
 Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
 Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
 Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
 Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
 Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
 Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
 Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
 Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
 Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
 Hernandez,J., Hernandez,O., Hodgson,A., Hoques,M., Holloway,C.,
 Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
 Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
 Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
 Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
 Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
 Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
 Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
 Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
 Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
 Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
 Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenwo,S.,
 Ogih,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
 Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
 Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M.,
 Ruiz,S., Savary,G., Scherret,S., Scott,G., Shen,H., Shoohtari,N.,
 Sison,I., Sodergren,E., Sonaik,T., Sparks,A., Stanley,H.,
 Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
 Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
 Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
 Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
 Watlington,S., Williams,G., Williamson,A., Wleczky,R., Wooden,S.,
 Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
 Weinstein,G. and Gibbs,R.

Direct Submission
 Unpublished
 2 (bases 1 to 31938)
 Worley,K.C.
 Direct Submission
 Submitted (15-SEP-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

TITLE
 JOURNAL
 REFERENCE
 2 (bases 1 to 31938)
 Worley,K.C.
 Direct Submission
 Submitted (15-SEP-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

On Dec 20, 2001 this sequence version replaced gi:17062140.
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GBCW
 Center clone name: CH230-5B20
 ----- Summary Statistics
 Assembly program: Phrap; version 0.990329First call to
 findPhrapList
 Consensus quality: 26514 bases at least Q40
 Consensus quality: 29012 bases at least Q30
 Consensus quality: 31138 bases at least Q20
 Estimated insert size: 15885; sum-of-contigs estimation
 Quality coverage: 0x in Q20 bases; agarose-gel estimation
 Quality coverage: 0.2x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 20 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence.
 * as soon as it is available and the accession number will
 * be preserved.
 *
 * 1 1998: contig of 1998 bp in length
 * 2098: gap of unknown length
 * 4623: contig of 2525 bp in length
 * 4723: gap of unknown length
 * 4724: contig of 2122 bp in length
 * 6845: gap of unknown length
 * 6846: gap of unknown length
 * 6946: contig of 2123 bp in length
 * 9068: gap of unknown length
 * 9169: contig of 1766 bp in length
 * 10934: gap of unknown length
 * 11034: contig of 1622 bp in length
 * 12656: gap of unknown length
 * 12657: contig of 1599 bp in length
 * 12757: gap of unknown length
 * 14356: contig of 1105 bp in length
 * 14456: gap of unknown length
 * 15560: gap of unknown length
 * 15661: contig of 1360 bp in length
 * 17020: gap of unknown length
 * 17120: contig of 1164 bp in length
 * 18284: gap of unknown length
 * 18385: gap of unknown length
 * 19755: contig of 1270 bp in length
 * 19755: gap of unknown length
 * 21265: contig of 1511 bp in length
 * 21266: gap of unknown length
 * 21366: contig of 1427 bp in length
 * 22793: gap of unknown length
 * 22893: contig of 1742 bp in length
 * 24635: gap of unknown length
 * 24735: contig of 1060 bp in length
 * 25795: gap of unknown length
 * 25895: contig of 1042 bp in length
 * 26937: gap of unknown length
 * 27037: contig of 1438 bp in length
 * 28475: gap of unknown length
 * 28575: contig of 1137 bp in length
 * 29712: gap of unknown length
 * 29812: contig of 1001 bp in length
 * 30812: gap of unknown length
 * 30813: contig of 1026 bp in length.
 * 30913
 Location/Qualifiers
 1..31938
 /organism="Rattus norvegicus"
 /db_xref="taxon:10116"
 /clone="CH230-5B20"

BASE COUNT 7910 a 6842 c 7612 g 7622 t 1952 others
 ORIGIN
 Query Match 85.0%; Score 17; DB 2; Length 31938;
 Best Local Similarity 100.0%; Pred. No. 2e+03;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 4 cccggccgaagccggg 20
 |||||
 Db 24061 CCCGCCGAGCCGGG 24077
 |||||
 RESULT 10
 AC103767
 LOCUS
 DEFINITION Homo sapiens chromosome 8 clone RP11-1077M7 map 8, LOW-PASS
 SEQUENCE SAMPLING.
 AC103767
 AC103767.1 GI:17149652
 HTG: HTGS_PHASE0.
 SOURCE
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE
 AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 TITLE Homo sapiens chromosome 8, clone RP11-1077M7
 JOURNAL Unpublished
 REFERENCE
 2 (bases 1 to 64620)
 AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
 Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B.,
 Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
 Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
 Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
 Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
 Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
 Jones,B., Kearfott,A., Horton,L., Hulme,W., Illiev,I., Johnson,R.,
 Jones,C., Katat,A., Karatas,A., Kells,C., LaRoque,K.,
 Lamazares,R., Landers,T., Lehotzky,J., Levine,N., Liu,G.,
 MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
 McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
 McNeil,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C.,
 Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
 Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
 Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
 Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupbach,R.,
 Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
 Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
 Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (29-NOV-2001) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L21727
 Center clone name: 1077_M7

 * NOTE: This record contains 83 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows

* overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

1
 675: contig of 675 bp in length
 676 775: gap of 100 bp
 776 1433: contig of 658 bp in length
 1434 1533: gap of 100 bp
 1534 2238: contig of 705 bp in length
 2239 2338: gap of 100 bp
 2339 3016: contig of 678 bp in length
 3017 3116: gap of 100 bp
 3117 3772: contig of 656 bp in length
 3773 3872: gap of 100 bp
 3873 4508: contig of 636 bp in length
 4509 4608: gap of 100 bp
 4609 5270: contig of 662 bp in length
 5271 5370: gap of 100 bp
 5371 6038: contig of 668 bp in length
 6039 6138: gap of 100 bp
 6139 6837: contig of 699 bp in length
 6838 6937: gap of 100 bp
 6938 7613: contig of 676 bp in length
 7614 7713: gap of 100 bp
 7714 8402: contig of 689 bp in length
 8403 8502: gap of 100 bp
 8503 9179: contig of 677 bp in length
 9180 9279: gap of 100 bp
 9280 9942: contig of 663 bp in length
 9943 10042: gap of 100 bp
 10043 10647: contig of 605 bp in length
 10648 10747: gap of 100 bp
 10748 11391: contig of 644 bp in length
 11392 11491: gap of 100 bp
 11492 12174: contig of 683 bp in length
 12175 12274: gap of 100 bp
 12275 12967: contig of 693 bp in length
 12968 13067: gap of 100 bp
 13068 13749: contig of 682 bp in length
 13750 13849: gap of 100 bp
 13850 14558: contig of 709 bp in length
 14559 14658: gap of 100 bp
 14659 15347: contig of 689 bp in length
 15348 15447: gap of 100 bp
 15448 16153: contig of 706 bp in length
 16154 16253: gap of 100 bp
 16254 16948: contig of 695 bp in length
 16949 17048: gap of 100 bp
 17049 17740: contig of 692 bp in length
 17741 17840: gap of 100 bp
 17841 18523: contig of 683 bp in length
 18524 18623: gap of 100 bp
 18624 19286: contig of 663 bp in length
 19287 19386: gap of 100 bp
 19387 20058: contig of 672 bp in length
 20059 20158: gap of 100 bp
 20159 20821: contig of 663 bp in length
 20822 20921: gap of 100 bp
 20922 21619: contig of 698 bp in length
 21620 21719: gap of 100 bp
 21720 22423: contig of 704 bp in length
 22424 22523: gap of 100 bp
 22524 23226: contig of 703 bp in length
 23227 23326: gap of 100 bp
 23327 24014: contig of 688 bp in length
 24015 24114: gap of 100 bp
 24115 24806: contig of 692 bp in length
 24807 24906: gap of 100 bp
 24907 25598: contig of 692 bp in length
 25599 25698: gap of 100 bp
 25699 26385: contig of 687 bp in length
 26386 26485: gap of 100 bp

* 26486 27145: contig of 660 bp in length
 * 27146 27445: gap of 100 bp
 * 27446 27936: contig of 691 bp in length
 * 27937 28036: gap of 100 bp
 * 28037 28719: contig of 683 bp in length
 * 28720 28819: gap of 100 bp
 * 28820 29475: contig of 656 bp in length
 * 29476 29575: gap of 100 bp
 * 29576 30265: contig of 690 bp in length
 * 30266 30365: gap of 100 bp
 * 30366 31064: contig of 699 bp in length
 * 31065 31164: gap of 100 bp
 * 31165 31862: contig of 698 bp in length
 * 31863 31962: gap of 100 bp
 * 31963 32676: contig of 714 bp in length
 * 32677 32776: gap of 100 bp
 * 32777 33451: contig of 675 bp in length
 * 33452 33551: gap of 100 bp
 * 33552 34244: contig of 693 bp in length
 * 34245 34344: gap of 100 bp
 * 34345 35035: contig of 691 bp in length
 * 35036 35135: gap of 100 bp
 * 35136 35798: contig of 663 bp in length
 * 35799 35898: gap of 100 bp
 * 35899 36566: contig of 668 bp in length
 * 36567 36666: gap of 100 bp
 * 36667 37335: contig of 669 bp in length
 * 37336 37435: gap of 100 bp
 * 37436 38084: contig of 649 bp in length
 * 38085 38184: gap of 100 bp
 * 38185 38877: contig of 693 bp in length
 * 38878 38977: gap of 100 bp
 * 38978 39675: contig of 698 bp in length
 * 39676 39775: gap of 100 bp
 * 39776 40428: contig of 653 bp in length
 * 40429 40528: gap of 100 bp
 * 40529 41202: contig of 674 bp in length
 * 41203 41302: gap of 100 bp
 * 41303 41991: contig of 689 bp in length
 * 41992 42091: gap of 100 bp
 * 42092 42760: contig of 669 bp in length
 * 42761 42860: gap of 100 bp
 * 42861 43518: contig of 658 bp in length
 * 43519 43618: gap of 100 bp
 * 43619 44309: contig of 691 bp in length
 * 44310 44409: gap of 100 bp
 * 44410 45082: contig of 673 bp in length
 * 45083 45182: gap of 100 bp
 * 45183 45862: contig of 680 bp in length
 * 45863 45962: gap of 100 bp
 * 45963 46613: contig of 651 bp in length
 * 46614 46713: gap of 100 bp
 * 46714 47378: contig of 665 bp in length
 * 47379 47478: gap of 100 bp
 * 47479 48175: contig of 697 bp in length
 * 48176 48275: gap of 100 bp
 * 48276 48935: contig of 660 bp in length
 * 48936 49035: gap of 100 bp
 * 49036 49744: contig of 709 bp in length
 * 49745 49844: gap of 100 bp
 * 49845 50534: contig of 690 bp in length
 * 50535 50634: gap of 100 bp
 * 50635 51332: contig of 698 bp in length
 * 51333 51432: gap of 100 bp
 * 51433 52111: contig of 679 bp in length
 * 52112 52211: gap of 100 bp
 * 52212 52872: contig of 661 bp in length
 * 52873 52972: gap of 100 bp
 * 52973 53605: contig of 633 bp in length

Query Match 85.0%; Score 17; DB 2; Length 64620;
 Best Local Similarity 100.0%; Pred. No. 1.6e+03;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 ctaccggccgaagggc 17
    |||||||
Db 4449 CTACCGCGCGAAGGCC 4465

RESULT 11
LOCUS CTY16136 2142 bp DNA linear BCT 23-FEB-2001
DEFINITION Clostridium thermoaceticum enr gene.
ACCESSION Y16136
VERSION Y16136.1 GI:3402835
KEYWORDS 2-enoate reductase; enr gene.
SOURCE Moorella thermoacetica.
ORGANISM Bacteria; Firmicutes; Bacillus/Clostridium group;
Thermoanaerobacter group; Moorella group; Moorella.
REFERENCE 1 (bases 1 to 2142)
AUTHORS Rohdich, F., Wiese, A., Felcht, R., Simon, H. and Bacher, A.
TITLE Enolate reductases of Clostridia. CLONING, SEQUENCING, AND
EXPRESSION
JOURNAL J. Biol. Chem. 276 (8), 5779-5787 (2001)
PUBMED 11060310
REFERENCE 2 (bases 1 to 2142)
AUTHORS Rohdich, F.
TITLE Direct Submission
JOURNAL Submitted (08-JAN-1998) F. Rohdich, TU Muenchen, Institut fuer
Organische Chemie und Biochemie, Lichtenbergstr. 4, D- 85748
Garching, FRG
REMARK revised by [3]
REFERENCE 3 (bases 1 to 2142)
AUTHORS Rohdich, F.
TITLE Direct Submission
JOURNAL Submitted (16-JUL-1998) F. Rohdich, TU Muenchen, Institut fuer
Organische Chemie und Biochemie, Lichtenbergstr. 4, D- 85748
Garching, FRG
COMMENT On Aug 7, 1998 this sequence version replaced gi:2791351.
Related location: Y09960.
FEATURES
    source
    1..2142
    /organism="Moorella thermoacetica"
    /strain="DSM 1974"
    /db_xref="taxon:1525"
    32..37
    /gene="enr"
    32..2050
    /gene="enr"
    47..2050
    /gene="enr"
    /EC_number="1.3.1.31"
    /codon_start=1
    /transl_table=11
    /evidence=experimental
    /product="2-enoate reductase"
    /protein_id="CAA76082.1"
    /db_xref="GI:3402836"
    /db_xref="SPTREMBL:O52935"
    /translation="MVAVYTRLPKIGKIGKVEIKNKIAMTPMVGVLGATHDGCFSKRVV
    DYVERAKGGLTITSVTKVDNEIERFKAGVPVATANPLHFIATAGELTERVHAYG
    AEIAVEAGFDGVEIHAMHEGYLLDQFTIALFNRRGDKYGGALDELTFFIETVRAIKD
    RVGKDFPVYLRFSKNYIKDWRQGLFGENFQEKGRDVEEPFLAAAKILLEGAYDGEDA
    DAGSYDAWYWAHPVYQKHGCLYPLTORLKEVVKVPVIVAGRLIEPELAEALVKQQA
    DMIAGRLGLTDPYVWKNKVTGRSKNIRPGICHDGCLGRGFLGRPLSCVTNPACGRE
    FEYAIDRAPEAKQVMYVGGVAGMEAAARVPALRGHVSLEYKSDRLGSHVVEAAVPDF
    KADGRLLKWTNLTGLOVEIHLNOEVTPEFEEKNPVVVATGTSTPAIPDIPGVN
    KDKVTYDMLLLKRGKAGDRVVIIGGLVGCETALWLAQGGKQVTFIILLDLMRAGI
    PVPYNNRMRLDLLKMGVKNWLTTSVLEVTDDGVTLLIGNYQNSPLPDTVILAVGF
    GADQLYNALRDKIPNLYLIGDSREPRNLAGIWEYGVGRGI"
    misc_binding
    1212..1278
    /gene="enr"
    /bound_moiety="FAD"
    1590..1626
    misc_binding

```

```

BASE COUNT 470 a 569 c 687 g 416 t
ORIGIN
/ gene="enr"
/ bound_moiety="NAD"

Query Match 84.0%; Score 16.8; DB 1; Length 2142;
Best Local Similarity 90.0%; Pred. No. 5.7e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctaccggccgaagggcggg 20
    |||||||
Db 1223 CTCCCGCGCGCAGCGCGGG 1204

RESULT 12
LOCUS AX078430 5451 bp DNA linear PAT 22-FEB-2001
DEFINITION Sequence 5 from Patent WO0107589.
ACCESSION AX078430
VERSION AX078430.1 GI:13158086
KEYWORDS Pseudomonas sp.
SOURCE Pseudomonas sp.
ORGANISM Bacteria; Proteobacteria.
REFERENCE 1 (bases 1 to 5451)
AUTHORS Fraser, C.M., Tuemmler, B., Hoheisel, J., Duesterhoeft, A., Hilbert, H.,
Timmis, K.N., Moore, E., Straetz, M., Heim, S. and Golyshin, P.
TITLE Dna sequences encoding lipopeptide synthases
JOURNAL Patent: WO 0107589-A 5 01-FEB-2001;
THE INSTITUTE FOR GENOMIC RESEARCH (US); OIAGEN GmbH (DE);
Gesellschaft fuer Biotechnologische Forschung mbH (GBF) (DE);
Deutsches Krebsforschungszentrum Stiftung des Oeffentlichen Rechts
(DE); Medizinische Hochschule Hannover (DE)
FEATURES
    source
    1..5451
    /organism="Pseudomonas sp."
    /db_xref="taxon:306"
    937 a 1766 c 1807 g 941 t
BASE COUNT 937 a 1766 c 1807 g 941 t
ORIGIN

Query Match 84.0%; Score 16.8; DB 6; Length 5451;
Best Local Similarity 90.0%; Pred. No. 4.2e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctaccggccgaagggcggg 20
    |||||||
Db 2049 CTACCGCGCGAAGCGCTGG 2068

RESULT 13
LOCUS AB019552 6220 bp DNA linear BCT 18-JAN-2002
DEFINITION Aeropyrum pernix genes for 16S rRNA, ITS, 23S rRNA, partial and
complete sequence, strain:OH2.
ACCESSION AB019552
VERSION AB019552.2 GI:18249697
KEYWORDS Aeropyrum pernix (strain:OH2) DNA.
SOURCE Aeropyrum pernix
ORGANISM Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
Desulfurococcaceae; Aeropyrum.
REFERENCE 1 (sites)
AUTHORS Nomura, N., Sako, Y., Morinaga, Y., Kogishi, T. and Uchida, A.
TITLE Intraspecific genetic polymorphism in the rRNA gene locus of the
hyperthermophilic archaeon Aeropyrum pernix, implying the presence
of multiple hotspots for intron homing
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 6220)
AUTHORS Nomura, N.
TITLE Direct Submission
JOURNAL Submitted (05-NOV-1998) Norimichi Nomura, Kyoto University, Lab. of

```

Marine Microbiology, Graduate School of Agriculture,
Kitashirakawa-Oiwake-cho, Sakyo-ku, Kyoto 606-8502, Japan
(E-mail: j54718@sakura.kudpc.kyoto-u.ac.jp, Tel: 81-75-753-6219,
Fax: 81-75-753-6226)

COMMENT On Jan 18, 2002 this sequence version replaced gi:3869179.

FEATURES

source
Location/Qualifiers

1..6220

/organism="Aeropyrum pernix"

/strain="OH2"

/db_xref="taxon:56636"

298..1798

/product="16S ribosomal RNA"

1799..2155

/note="internal transcribed spacer (ITS) region"

2156..5238

/product="23S ribosomal RNA"

1088 a 1878 c 2370 g 884 t

BASE COUNT

ORIGIN

Query Match 84.0%; Score 16.8; DB 1; Length 6220;

Best Local Similarity 90.0%; Pred. No. 4e+03;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ctaccggcggaagcgcg 20

||||| ||||| ||||| |||||

Db 3812 CTAACCTGCCGAGCGGG 3831

RESULT 14

AB078015

LOCUS

DEFINITION Aeropyrum pernix genes for 16S rRNA, 23S rRNA and ITS region, BCT 19-JAN-2002

complete sequence, strain:OH1.

AB078015

ACCESSION

VERSION AB078015.1 GI:18250956

KEYWORDS

SOURCE Aeropyrum pernix (strain:OH1) DNA.

ORGANISM

Aeropyrum pernix

Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;

1 (sites)

Nomura, N., Morinaga, Y., Kogishi, T., Kim, E., Sako, Y. and Uchida, A.

Heterogeneous yet partially similar introns reside in identical

positions of the rRNA genes in natural isolates of the archaeon

Aeropyrum pernix

Unpublished

2 (bases 1 to 6220)

Nomura, N.

Direct Submission

Submitted (16-JAN-2002) Norimichi Nomura, Kyoto University,

Division of Applied Biosciences, Graduate School of Agriculture,

Sakyo, Kyoto 606-8502, Japan

(E-mail: j54718@sakura.kudpc.kyoto-u.ac.jp, Tel: 81-75-753-6224,

Fax: 81-75-753-6226)

FEATURES

source

1..6220

/organism="Aeropyrum pernix"

/strain="OH1"

/db_xref="taxon:56636"

298..1797

/gene="arnS"

/product="16S ribosomal RNA"

298..1797

/gene="arnS"

1798..2155

/note="internal transcribed spacer (ITS) region"

2156..5238

/gene="arnL"

/product="23S ribosomal RNA"

2156..5238

/gene="arnL"

1086 a 1877 c 2372 g 885 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity 84.0%; Score 16.8; DB 1; Length 6220;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ctaccggcggaagcgcg 20

||||| ||||| ||||| |||||

Db 3812 CTAACCTGCCGAGCGGG 3831

RESULT 15

AB078016

LOCUS

DEFINITION Aeropyrum pernix genes for 16S rRNA, 23S rRNA and ITS region, BCT 19-JAN-2002

complete sequence, strain:OH3.

AB078016

ACCESSION

VERSION AB078016

KEYWORDS

SOURCE Aeropyrum pernix (strain:OH3) DNA.

ORGANISM

Aeropyrum pernix

Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;

1 (sites)

Nomura, N., Morinaga, Y., Kogishi, T., Kim, E., Sako, Y. and Uchida, A.

Heterogeneous yet partially similar introns reside in identical

positions of the rRNA genes in natural isolates of the archaeon

Aeropyrum pernix

Unpublished

2 (bases 1 to 6220)

Nomura, N.

Direct Submission

Submitted (16-JAN-2002) Norimichi Nomura, Kyoto University,

Division of Applied Biosciences, Graduate School of Agriculture,

Sakyo, Kyoto 606-8502, Japan

(E-mail: j54718@sakura.kudpc.kyoto-u.ac.jp, Tel: 81-75-753-6224,

Fax: 81-75-753-6226)

FEATURES

source

1..6220

/organism="Aeropyrum pernix"

/strain="OH3"

/db_xref="taxon:56636"

298..1798

/gene="arnS"

/product="16S ribosomal RNA"

298..1798

/gene="arnS"

1799..2155

/note="internal transcribed spacer (ITS) region"

2156..5238

/gene="arnL"

/product="23S ribosomal RNA"

2156..5238

/gene="arnL"

1091 a 1877 c 2368 g 884 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity 84.0%; Score 16.8; DB 1; Length 6220;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ctaccggcggaagcgcg 20

||||| ||||| ||||| |||||

Db 3812 CTAACCTGCCGAGCGGG 3831

Search completed: June 19, 2002, 15:43:19

Job time: 13457 sec

Result No.	Score	Query Match	Length	DB ID	Description
1	20	100.0	44	20	AAFX02689
2	20	100.0	44	20	AAFX02687
3	16.8	84.0	5451	22	AAFP26319
4	16	80.0	341	22	AAH68825
5	15.8	79.0	2040	22	AAH52030
6	15.8	79.0	2388	22	AAAD16357
7	15.8	79.0	2388	22	AAAI66145
8	15.8	79.0	3217	22	AAAI66138
9	15.8	79.0	3226	12	AAQ11038
					CFDV-DNA fragment
					CFDV-DNA fragment
					Pseudomonas sp lip
					Human cervical can
					Mycobacterium tube
					Human sbq34976IGBA
					PAM polynucleotide
					PAM encoding cDNA
					Encodes Rat C-term

```
FT FT /*tag= f
XX PN /note= "Region binds to nucleotides 27 to 32"
XX PN
XX PD
XX XX
XX PF 16-JUL-1997; 97DE-1030502.
XX PR 16-JUL-1997; 97DE-1030502.
XX XX
XX PA (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX PI Becker D, Hehn A, Randles JW, Rohde W, Salamini F;
XX DR WPI; 1999-096863/09.
XX XX
XX XX Coconut foliar decay virus promoters - for gene expression in
XX PT bacteria and yeasts
XX PS Disclosure; Fig 4; 14pp; German.
XX CC This invention describes a coconut foliar decay virus (CFDV) DNA fragment
XX CC that includes the stem-loop structure of CFDV DNA but lacks the
XX CC translation start codons of open reading frames ORF1 and/or ORF2. The
XX CC new CFDV DNA fragment is useful as a bacterial or yeast promoter, as a
XX CC promoter for tissue-specific (especially phloem-specific) gene expression
XX CC in plants and for production of chimeric constructs for transient or
XX CC stable expression. Certain fragments of CFDV DNA have stronger promoter
XX CC activity in E. coli than the CamV 35S promoter.
XX XX
XX SQ Sequence 44 BP; 8 A; 16 C; 17 G; 3 T; 0 other;

Query Match 100.0%; Score 20; DB 20; Length 44;
Best Local Similarity 100.0%; Pred. No. 6.3;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctaccggcggaagccggg 20
DB 24 ctaccggcggaagccggg 43

RESULT 2
AA02687
ID AAX02687 standard; DNA; 44 BP.
XX AC AAX02687;
XX DT 10-MAY-1999 (first entry)
XX DE CFDV-DNA fragment stem loop repeat region.
XX KW Stem loop; coconut foliar decay virus; CFDV; bacterial promoter;
XX KW yeast promoter; tissue-specific; phloem-specific; CamV 35S promoter; ss.
XX OS Coconut foliar decay virus
XX FH Key Location/Qualifiers
FT stem_loop 4..22
FT /*tag= a
FT /*note= "RPT1"
FT misc_binding 4..10
FT /*tag= b
FT /*note= "Region binds to nucleotides 16 to 22"
FT misc_binding 16..22
FT /*tag= c
FT /*note= "Region binds to nucleotides 4 to 10"
FT stem_loop 27..43
FT /*tag= d
FT /*note= "RPT2"
FT misc_binding 27..32
FT /*tag= e

/Note= "Region binds to nucleotides 38 to 43"
/*tag= f
/*note= "Region binds to nucleotides 27 to 32"
DE19730535-A1.
21-JAN-1999.
16-JUL-1997; 97DE-1030535.
16-JUL-1997; 97DE-1030535.
(PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
Becker D, Hehn A, Randles JW, Rohde W, Salamini F;
WPI; 1999-096867/09.
Coconut foliar decay virus promoters - for gene expression in
plants, bacteria and yeasts
Disclosure; Fig 4; 12pp; German.
This invention describes a coconut foliar decay virus (CFDV) DNA fragment
that includes the stem-loop structure of CFDV DNA but lacks the
translation start codons of open reading frames ORF1 and/or ORF2. The
new CFDV DNA fragment is useful as a bacterial or yeast promoter, as a
promoter for tissue-specific (especially phloem-specific) gene expression
in plants and for production of chimeric constructs for transient or
stable expression. Certain fragments of CFDV DNA have stronger promoter
activity in E. coli than the CamV 35S promoter.
Sequence 44 BP; 8 A; 16 C; 17 G; 3 T; 0 other;

Query Match 100.0%; Score 20; DB 20; Length 44;
Best Local Similarity 100.0%; Pred. No. 6.3;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctaccggcggaagccggg 20
DB 24 ctaccggcggaagccggg 43

RESULT 3
AAF26319
ID AAF26319 standard; DNA; 5451 BP.
XX AC AAF26319;
XX DT 02-MAY-2001 (first entry)
XX DE Pseudomonas sp lipopeptide synthase DNA ORF08563.
XX KW Lipopeptide synthase; transgenic plant; antiviral; antibacterial;
XX KW antifungal; surfactant; microbial depollution; emulsifier;
XX KW tertiary crude oil recovery; ORF08563; ds.
XX OS Pseudomonas sp.
XX PN WO200107589-A2.
XX PD 01-FEB-2001.
XX PF 21-JUL-2000; 2000WO-EP07002.
XX PR 27-JUL-1999; 99DE-1035106.
XX XX (TIGR-) TIGR INST GENOMIC RES.
XX PA (QIAG-) QIAGEN GMBH.
XX PA (GBF-) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.
XX PA (DKFZ-) DKFZ DEUT KREBSFORSCHUNGSZENTRUM.
```

(MEDI-) MEDIZINISCHE HOCHSCHULE HANNOVER.

Fraser CM, Tuemmler B, Hohelsel J, Diesterhoeft A, Hilbert H;
Timmis KN, Moore E, Straetz M, Heim S, Golyshin P;
WPI: 2001-159715/16.

New DNA encoding a bacterial lipopeptide synthase, useful e.g. for
producing antimicrobial agents, surfactants with increased
bioavailability in microbial depollution methods, and transgenic plants

Claim 3a; Page 28-30; 33pp; German.

This invention describes a novel DNA sequence (I) that expresses products
having the biological function of lipopeptide synthase. The invention
also describes (1) recombinant expression vectors containing (I); (2)
prokaryotic and eukaryotic cells transformed or transfected with (I) or
the vector of (2); (3) production of lipopeptide synthases by culturing
cells of (2); (4) expression products (II) of (I), and synthetic proteins
or peptides with the same sequences; (5) mono- or poly-clonal antibodies
(Ab) specific for (II); (6) hybridoma cells that produce monoclonal Ab;
and (7) transgenic plants that contain cells of (2), (I), and their
fragments, are useful for expression of recombinant lipopeptide
synthases, and as probes and primers for detection, isolation and
amplification of full-length cDNA sequences. (I) are used to produce
transgenic plants. Lipopeptide synthases are useful for production of
antiviral, antibacterial or antifungal lipopeptides, particularly useful
for treatment of plants, and for production of lipopeptide surfactants
used for increasing bioavailability in microbial depollution processes
(acting as emulsifiers) and in tertiary crude oil recovery. Individual
subunits from different lipopeptide synthases may be combined to allow
synthesis of many different biologically active substances.

Sequence 5451 BP; 937 A; 1766 C; 1807 G; 941 T; 0 other;

Query Match 84.0%; Score 16.8; DB 22; Length 5451;
Best Local Similarity 90.0%; Pred. No. 1.2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ctaccgcgcgaagccggg 20
|||||
Db 2049 ctaccgcgcgaagccgtgg 2068

RESULT 4
AAH68825
ID AAH68825 standard; CDNA; 341 BP.

XX AC AAH68825;
XX DT 19-SEP-2001 (first entry)
XX DE Human cervical cancer marker nucleic acid 99.

XX KW Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.
XX OS Homo sapiens.
XX PN WO20012467-A2.

XX PD 14-JUN-2001.

XX PF 08-DEC-2000; 2000WO-US33312.
XX PR 08-DEC-1999; 99US-0169681.
XX PR 21-DEC-1999; 99US-0171350.
XX PR 14-MAR-2000; 2000US-0189315.
XX PR 12-MAY-2000; 2000US-0203791.
XX PR 09-JUN-2000; 2000US-0210600.
XX PR 21-JUL-2000; 2000US-0220114.

(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

Schlegel R, Deeds J, Berger A, Zhao X;
WPI: 2001-375006/39.

New isolated nucleic acid for diagnosing and treating cervical cancer
and for assessing and detecting compounds for treating the cancer -
Claim 1; Page 120; 1051pp; English.

The invention relates to novel genes (AAH68727-AAH73383) associated with
cervical cancer with cytostatic activity. The nucleic acids and encoded
polypeptides are useful: to assess if a patient is afflicted with
cervical cancer or has a pre-malignant condition; to monitor the
progression of cervical cancer or a premalignant condition in a patient;
and to select and/or assess the efficacy of a compound or therapy for
inhibiting cervical cancer in a patient. The nucleic acids may also be
useful for gene therapy.

Sequence 341 BP; 72 A; 93 C; 93 G; 68 T; 15 other;

Query Match 80.0%; Score 16; DB 22; Length 341;
Best Local Similarity 94.1%; Pred. No. 3.2e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 ccgcgcgcgaagccggg 20
|||||
Db 192 ccgcgcgcgaagccggg 208

RESULT 5
AAH52030
ID AAH52030 standard; DNA; 2040 BP.

XX AC AAH52030;
XX DT 04-SEP-2001 (first entry)

XX DE Mycobacterium tuberculosis potential drug target gene SEQ ID 84.
XX KW Drug target; growth; organism viability; characterisation; ds.
XX OS Mycobacterium tuberculosis.

XX PN WO200135317-A1.

XX PD 17-MAY-2001.

XX PF 13-NOV-2000; 2000WO-US31152.

XX PR 12-NOV-1999; 99US-0165086.

XX PR 12-NOV-1999; 99US-0165124.

XX PR 01-FEB-2000; 2000US-0179531.

XX PA (REGC) UNIV CALIFORNIA.

XX PI Eisenberg D, Rotstein SH, Marcotte EM;

XX DR WPI: 2001-329193/34.

XX DR P-PSDB; AAG81179.

XX PT Identifying nucleotide or polypeptide sequence for use as drug target,
XX PT involves providing algorithm that analyzes a functional relationship
XX PT between nucleotide or polypeptide sequences, and comparing the
XX PT sequences -

XX PS Disclosure; Page 103-104; 207pp; English.

XX CC This invention relates to a method for identifying a nucleotide or
XX CC polypeptide sequence that may be a drug target, or essential for growth
XX CC or viability of an organism. Polynucleotide sequences AAH51947 - AAH52092

CC represent DNA encoding proteins AAG81096 - AAG81241, Mycobacterium
 CC tuberculosis proteins which are potential drug targets. The DNA and
 CC protein sequences are used to illustrate the method of the invention. The
 CC method involves providing an unknown nucleotide or polypeptide sequences,
 CC and comparing it to a number of sequences along with at least one
 CC algorithm capable of analysing a functional relationship between
 CC nucleotide and polypeptide sequences. The method is useful for
 CC characterising the function of nucleic acids and polypeptides that may be
 CC useful as a target for a drug or essential for the growth or viability of
 CC an organism.
 XX
 SQ Sequence 2040 BP; 374 A; 673 C; 671 G; 322 T; 0 other;

Query Match 79.0%; Score 15.8; DB 22; Length 2040;
 Best Local Similarity 89.5%; Pred. No. 3.5e+02;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 taccggcggaagcgcg 20
 |||| ||||| |||||
 Db 1731 taccggcggaagcgcg 1749

RESULT 6
 AAD16357/C
 ID AAD16357 standard; DNA; 2238 BP.

XX AC AAD16357;

XX DT 19-NOV-2001 (first entry)

XX DE Human sbg34976IGBa gene #1.

XX Human; Alzheimer's disease; anyotropic lateral sclerosis;
 KW ALS; Zollinger-Ellison syndrome; immune system disease; schizophrenia;
 KW inflammation; haematopoietic disease; anxiety; feeding disorder; aging;
 KW anorexia; depression; cardiovascular disease; sleep disorder; seizure;
 KW memory alteration; migraine; stroke; asthma; neuropathy; hypoglycaemia;
 KW sexual disorder; growth abnormality; infection; autoimmune disease;
 KW rheumatoid arthritis; cataractogenesis; angiogenesis; atherosclerosis;
 KW cerebral ischaemia; cirrhosis; Huntington's disease; Hodgson's disease;
 KW hypercholesterolaemia; headache; amnesia; cardiac arrhythmia; obesity;
 KW diabetes mellitus; glomerulonephritis; renovascular hypertension;
 KW cancer; vaccine; gene therapy; sbg34976IGBa gene; ds.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers
 XX CDS 1..2238
 XX FT /*tag= a
 XX FT /product= "Human sbg34976IGBa protein #1"

XX WO200160850-A1.

XX PD 23-AUG-2001.

XX PF 14-FEB-2001; 2001WO-US04703.

XX PR 14-FEB-2000; 2000US-0182172.

XX PR 29-FEB-2000; 2000US-0185084.

XX PR 18-APR-2000; 2000US-0198583.

XX PR 04-OCT-2000; 2000US-0237963.

XX (SMIK) SMITHKLINE BEECHAM CORP.

XX (SMIK) SMITHKLINE BEECHAM PLC.

XX PI Agarwal P, Kabnick KS, Murdock PR, Rizvi SK, Smith RF, Xiang Z;

XX WPI; 2001-536566/59.

XX P-PSDB; AAE09450.

XX New secreted and membrane associated polypeptides for treating
 PT Alzheimer's disease, psoriasis, cancer, enterocolitis, sleep and sexual

PT disorders, stroke, and asthma
 XX
 PS Claim 2; Page 48-49; 94pp; English.

XX The present sequence is a gene encoding human sbg34976IGBa protein,
 CC a membrane bound protein of the invention.
 CC The invention relates to secreted and membrane associated polypeptides
 CC and nucleic acid molecules encoding such polypeptides. Sequences of the
 CC invention are useful for treating diseases such as Alzheimer's disease,
 CC amyotrophic lateral sclerosis (ALS), Zollinger-Ellison syndrome, diseases
 CC of the immune system, haematopoietic disease, inflammation, anxiety,
 CC schizophrenia, feeding disorders, anorexia, depression, social, sexual
 CC and rewarded behaviour, cardiovascular disease, sleep disorder, learning
 CC and memory alteration and altered immune response, seizure, migraine,
 CC cancer, stroke, asthma, neuropathy, aging, sexual disorders, treatment
 CC of transsexuals, growth abnormalities, obesity, infections, autoimmune
 CC diseases (e.g. rheumatoid arthritis), cataractogenesis, angiogenesis,
 CC disorders associated with healthy maintenance of gastric mucosa and
 CC repair of acute and chronic mucosal lesion, lung carcinoma, cerebral
 CC ischaemia, atherosclerosis, cirrhosis, Huntington's disease, headache,
 CC amnesia, multiple sclerosis, Hodgson's disease, dilated cardiomyopathy,
 CC congestive heart failure, cardiac arrhythmias, hypercholesterolaemia,
 CC viral and non-viral hepatitis, type I and type II diabetes mellitus,
 CC glomerulonephritis, renovascular hypertension, hypoglycaemia, periodic
 CC paralyse, tendinitis and malignant hyperthermia. Polypeptides of the
 CC invention are used to identify membrane bound and soluble receptors.
 CC They are also useful as vaccines for inducing an immunological response
 CC in a mammal. Polynucleotides of the invention are used in gene therapy.
 CC They are also valuable for chromosome localisation studies and tissue
 CC expression studies.

XX Sequence 2238 BP; 400 A; 744 C; 743 G; 351 T; 0 other;

Query Match 79.0%; Score 15.8; DB 22; Length 2238;
 Best Local Similarity 89.5%; Pred. No. 3.4e+02;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 taccggcggaagcgcg 20
 ||||| ||||| |||||
 Db 1684 TACCGCGCGCGAGCGCG 1666

RESULT 7
 AAI66145/C
 ID AAI66145 standard; cDNA; 2388 BP.

XX AC AAI66145;

XX DT 14-JAN-2002 (first entry)

XX DE PAM polynucleotide 3.

XX Chinese hamster ovary; PAM; ECI.14.17.3; ss;
 KW Peptidylglycine alpha-amidating monooxygenase.

XX OS Unidentified.

XX CNL167154-A.

XX PD 10-DEC-1997.

XX PF 16-APR-1997; 97CN-0106382.

XX PR 16-APR-1997; 97CN-0106382.

XX (SHAN-) SHANGHAI BIO-CHEM INST CHINESE ACAD SCI.

XX Li B, Jiang Z, Xia O;

XX WPI; 2001-590349/67.

XX Amidase gene and modifying application of its expressed active products

PT to polypeptide amidating -

PS Example 1; Fig 2; 37pp; Chinese.

XX The invention relates to a rat amidase gene, peptidylglycine

CC alpha-amidating monooxygenase/PAM (EC1.14.17.3) and its active expression
CC product in Chinese hamster ovary (CHO) cell. The invention uses in situ
CC hybridization and a polymerase chain reaction (PCR) method to screen and
CC obtain three rPAM gene fragments, then through technologies of point
CC mutation and PCR recombination to splice rPAM holoenzyme gene and form
CC its eucaryon expression plasmid so as to obtain the active expression in
CC CHO cell. The obtained expression product can be directly used for
CC amidation modification of polypeptide and protein. The present sequence
CC is that of a polynucleotide of the PAM gene (AAI66138), useful to the
XX invention.

SQ Sequence 2388 BP; 638 A; 549 C; 589 G; 612 T; 0 other;

Query Match 79.0%; Score 15.8; DB 22; Length 2388;
Best Local Similarity 89.5%; Pred. No. 3.4e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctaccggcggaagggccgg 19
||||||| ||||| ||

Db 1823 CTACCGGGCGAAGGCTGG 1805

RESULT 8

AAI66138/C
ID AAI66138 standard; cDNA; 3217 BP.

XX AC AAI66138;

XX DT 14-JAN-2002 (first entry)

XX DE PAM encoding cDNA.

XX KW Chinese hamster ovary; PAM;

XX KW Peptidylglycine alpha-amidating monooxygenase; ss.

XX OS Unidentified.

XX FH Key Location/Qualifiers

XX CDS 11..2629

XX FT /*tag= a

XX FT /product= "PAM"

XX FT /EC_number= "1.14.17.3"

XX PN CNL167154-A.

XX PD 10-DEC-1997.

XX PF 16-APR-1997; 97CN-0106382.

XX PR 16-APR-1997; 97CN-0106382.

XX PA (SHAN-) SHANGHAI BIO-CHEM INST CHINESE ACAD SCI.

XX PI Li B, Jiang Z, Xia Q;

XX DR WPI; 2001-590349/67.

XX DR P-PSDB; AAM51684.

XX PT Amidase gene and modifying application of its expressed active products

XX PT to polypeptide amidating -

XX PS Claim 2; Fig 1; 37pp; Chinese.

XX CC The invention relates to a rat amidase gene, peptidylglycine

CC alpha-amidating monooxygenase/PAM (EC1.14.17.3) and its active expression

CC product in Chinese hamster ovary (CHO) cell. The invention uses in situ

CC hybridization and a polymerase chain reaction (PCR) method to screen and

CC obtain three rPAM gene fragments, then through technologies of point
CC mutation and PCR recombination to splice rPAM holoenzyme gene and form
CC its eucaryon expression plasmid so as to obtain the active expression in
CC CHO cell. The obtained expression product can be directly used for
CC amidation modification of polypeptide and protein.

SQ Sequence 3217 BP; 839 A; 762 C; 791 G; 825 T; 0 other;

Query Match 79.0%; Score 15.8; DB 22; Length 3217;

Best Local Similarity 89.5%; Pred. No. 3.4e+02;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctaccggcggaagggccgg 19
||||||| ||||| ||

Db 2652 CTACCGGGCGAAGGCTGG 2634

RESULT 9

AAQ11038/C

ID AAQ11038 standard; cDNA; 3226 BP.

XX AC AAQ11038;

XX DT 17-MAY-1991 (first entry)

XX DE Encodes Rat C-terminal amidating enzyme.

XX KW C-terminal amidating enzyme; calcitonin; gastrin; ss.

XX OS Rattus.

XX FH Key Location/Qualifiers

XX CDS 2..2563

XX FT /*tag= a

XX FT /product= C-terminal amidating enzyme

XX PN WO9102790-A.

XX PD 07-MAR-1991.

XX PF 14-AUG-1990; 90WO-JP01036.

XX PR 15-AUG-1989; 89JP-0209687.

XX PR 31-OCT-1989; 89JP-0281933.

XX PR 26-MAR-1990; 90JP-0076331.

XX PR 24-APR-1990; 90JP-0106412.

XX PR 02-AUG-1990; 90JP-0205475.

XX PA (SHIS) SHISEIDO KK.

XX WPI; 1991-087274/12.

XX DR P-PSDB; AAR11110.

XX PT New enzymes useful in peptide C-amidation - convert peptide C-
XX terminal glycine to alpha hydroxy glycine then remove this leaving
XX peptide C-term amide.

XX PS Disclosure; Fig 6; 114pp; Japanese.

XX CC The enzyme encoded by this sequence is involved in the C-terminal
XX amidation reactions. It can be used to amide physiologically
XX active peptides such as calcitonin and gastrin. It also amidates D-
XX Tyr-Trp-Gly, Phe-Gly-Phe-Gly and Gly-Phe-Gly. The region from
XX position 2 to 630 was used as a probe to isolate the C-terminal
XX amidating enzyme from the horse.
XX See also AAQ11039-Q11042.

SQ Sequence 3226 BP; 845 A; 748 C; 788 G; 844 T; 1 other;

Query Match 79.0%;

Best Local Similarity 89.5%; Pred. No. 3.3e+02; Length 3226;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ctaccgccggaagccgg 19
 ||||| ||||| ||

Db 2585 CTACCCGGCGAGGCTGG 2567

RESULT 10
 AAQ05637/c
 ID AAQ05637 standard; DNA; 3586 BP.

XX AC AAQ05637;

XX DT 14-DEC-1990 (first entry)

XX DE Sequence which hybridises with DNA encoding alpha amidating enzyme.

XX KW Alpha amidating enzyme; peptidyl amide; ss.

XX FH Key Location/Qualifiers
 FT CDS 346..2907
 FT /*tag= a

XX PN EP382403-A.

XX PD 16-AUG-1990.

XX PF 01-FEB-1990; 90EP-0301034.

XX PR 06-FEB-1989; 89US-0307366.

XX PA (UNIG-) UNIGENE LAB INC.

XX PI Betelsen AH, Mehta NM, Beaudry GA;

XX DR WPI; 1990-248308/33.
 DR P-PSDB; AAR96426.

XX PT Prodn. of alpha amidating enzyme - using prokaryotic or
 PT eukaryotic expression vectors having transcriptional promoter

XX PS Claim 2; page 17; 24pp; English.

XX CC This sequence hybridises, under stringent conditions, with a DNA
 CC sequence (I) which encodes an alpha amidating enzyme (AEE) and is
 CC foreign to the host cell into which it is transformed. (I)
 CC is connected to a promoter and is contained in an expression
 CC vector. The AEE encoded is used as a catalyst in the conversion
 CC of a peptidyl substrate to a corresp. peptidyl amide. This can be
 CC used for making a protein biologically active, e.g. calcitonin or
 CC growth hormone releasing factor. See also AAQ05638.

XX SQ Sequence 3586 BP; 908 A; 879 C; 906 G; 893 T; 0 other;

Query Match 79.0%; Score 15.8; DB 11; Length 3586;
 Best Local Similarity 89.5%; Pred. NO. 3.3e+02;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ctaccgccggaagccgg 19
 ||||| ||||| ||

Db 2929 CTACCCGGCGAGGCTGG 2911

RESULT 11
 AAQ10277/c
 ID AAQ10277 standard; DNA; 3586 BP.

XX AC AAQ10277;

XX DT 08-APR-1991 (first entry)

XX DE Type A alpha-amidating enzyme.

XX KW Amidation; post-translational modification; AE; ss.

XX OS Rattus rattus.

XX FH Key Location/Qualifiers
 FT CDS 346..2907
 FT /*tag= a
 FT /label= AE Type A

XX PN AU9049043-A.

XX PD 29-NOV-1990.

XX PF 01-FEB-1990; 90AU-0049043.

XX PR 06-FEB-1989; 89US-0307336.

XX PA (UNIG-) UNIGENE LAB INC.

XX PI Betelsen AH, Mehta NM, Beaudry GA;

XX DR WPI; 1991-022433/04.
 DR P-PSDB; AAR10322.

XX PT Expression systems for amidating enzyme - comprises prokaryotic
 PT or eukaryotic hosts contg. a recombinant expression vector contg.
 PT the corresp. DNA sequence.

XX PS Disclosure; Page 2; 52pp; English.

XX CC The DNA was isolated from libraries prepd. from RNA extracted from
 CC rat medullary thyroid carcinoma tissues or derived cell lines such
 CC as IVI 10028. The cDNAs extracted could be divided into distinct
 CC types, the sequence below being of Type A. (Type B is given in
 CC AAQ10278). The sequence can be inserted into expression vectors for
 CC the prodn. of recombinant alpha-amidating enzyme (AE) useful for
 CC post translational modification of other recombinant polypeptides
 CC such as calcitonin. The sequence includes a membrane spanning
 CC domain which may be undesirable in a recombinant DNA expression
 CC system, possibly inactivating the enzyme. A stop codon can be
 CC placed upstream of the domain-encoding sequence, pref. between
 CC posns. 2025 and 2275 (counting from the CDS). This results in the
 CC expression of a 75 rather than 94 kD. The sequence of Type A and
 CC Type B are substantially identical with the exception of an intron
 CC region from bases 1178-1492 (CDS) of the Type B cDNA.

XX SQ Sequence 3586 BP; 909 A; 879 C; 906 G; 892 T; 0 other;

Query Match 79.0%; Score 15.8; DB 12; Length 3586;
 Best Local Similarity 89.5%; Pred. NO. 3.3e+02;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ctaccgccggaagccgg 19
 ||||| ||||| ||

Db 2929 CTACCCGGCGAGGCTGG 2911

RESULT 12
 AAS21277/c
 ID AAS21277 standard; cDNA; 4185 BP.

XX AC AAS21277;

XX DT 24-OCT-2001 (first entry)

XX DE Human cDNA sequence encoding for PRO4329 polypeptide.

XX KW Human secretory and transmembrane; PRO; mammalian; cancer; lung;
 KW breast; prostate; cervical; tumour necrosis factor-alpha; TNF-alpha;
 KW cartilage; ear; proliferation; glucose; free fatty acid; skeletal muscle;
 KW adipocyte; A-peptide; factor VIIA; gene therapy; ss.

XX OS Homo sapiens.
XX PN WO200110466-A2.
XX PD 07-JUN-2001.
XX PF 01-DEC-2000; 2000WO-US32678.
XX PR 01-DEC-1999; 99WO-US28301.
XX PR 01-DEC-1999; 99WO-US28634.
XX PR 02-DEC-1999; 99WO-US28551.
XX PR 02-DEC-1999; 99WO-US28564.
XX PR 02-DEC-1999; 99WO-US28565.
XX PR 09-DEC-1999; 99US-0170262.
XX PR 16-DEC-1999; 99WO-US30095.
XX PR 20-DEC-1999; 99WO-US30911.
XX PR 20-DEC-1999; 99WO-US30999.
XX PR 30-DEC-1999; 99WO-US31243.
XX PR 06-JAN-2000; 2000WO-US00277.
XX PR 06-JAN-2000; 2000WO-US00376.
XX PR 11-FEB-2000; 2000WO-US03565.
XX PR 18-FEB-2000; 2000WO-US04341.
XX PR 18-FEB-2000; 2000WO-US04342.
XX PR 22-FEB-2000; 2000WO-US04414.
XX PR 24-FEB-2000; 2000WO-US04914.
XX PR 01-MAR-2000; 2000WO-US05004.
XX PR 20-MAR-2000; 2000WO-US05601.
XX PR 20-MAR-2000; 2000WO-US07377.
XX PR 21-MAR-2000; 2000WO-US07532.
XX PR 30-MAR-2000; 2000WO-US08439.
XX PR 17-MAY-2000; 2000WO-US13705.
XX PR 22-MAY-2000; 2000WO-US14042.
XX PR 30-MAY-2000; 2000WO-US14941.
XX PR 02-JUN-2000; 2000WO-US15264.
XX PR 10-NOV-2000; 2000WO-US30873.
XX PA (GETH) GENENTECH INC.
XX PI Baker KP, Beresini M, Deforge L, Desnoyers L, Filvaroff E, Gao W;
XX PI Gerritsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;
XX PI Smith V, Stewart TA, Tumas D, Watanabe CK, Wood WI, Zhang Z;
XX WPI: 2001-408281/43.
XX P-PSDB: AAU12205.
XX PT Isolated, secretory and transmembrane PRO polypeptide used to detect
XX PT other PRO polypeptides, link bioactive molecules to cells expressing
XX PT PRO polypeptides, and detect the presence of mammalian tumours e.g.
XX PT lung, breast, prostate, cervical -
XX PS Claim 3; Fig 67; 813pp; English.
XX CC AAS21244-AAS21518 encode for novel human secretory and transmembrane
XX CC PRO polypeptides. The PRO polypeptides are useful to detect other
XX CC PRO polypeptides, to link bioactive molecules to cells expressing
XX CC PRO polypeptides, to modulate biological activities of cells expressing
XX CC PRO polypeptides, and to detect the presence of mammalian lung, colon,
XX CC breast, prostate, rectal, cervical or liver tumours by comparing PRO
XX CC polypeptide expression in a cell sample to that in a control sample.
XX CC Some of the 275 sequences are also useful to stimulate the release of
XX CC tumour necrosis factor-alpha (TNF-alpha) from human blood, the
XX CC proliferation or differentiation of chondrocytes, the proliferation or
XX CC gene expression in pericyte cells, the release of proteoglycans from
XX CC cartilage, the proliferation of inner ear utricular supporting cells or
XX CC of T-lymphocytes, the release of a cytokine from peripheral blood
XX CC monocytes (PBMCs), or the proliferation of endothelial cells. Some of
XX CC the PRO polypeptides may modulate glucose or free fatty acid uptake by
XX CC skeletal muscle cells or by adipocytes; or inhibit binding of A-peptide
XX CC to factor VIIA. The PRO polypeptides can be used in assays to identify
XX CC molecules involved in binding interactions. The polynucleotides encoding
XX CC PRO polypeptides can be used to generate probes, antisense RNA/DNA,
XX CC transgenic or knock out animals and can be used in gene therapy.

XX SQ Sequence 4185 BP; 713 A; 1363 C; 1307 G; 802 T; 0 other;
XX
XX Query Match 79.0%; Score 15.8; DB 22; Length 4185;
XX Best Local Similarity 89.5%; Pred. No. 3.3e+02;
XX Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX QY 2 taccgcgcgcgaagccggg 20
XX ||||| |||||
XX Db 1820 TACCCGCGCGCAGCGCGG 1802
XX
XX RESULT 13
XX AAS59517/c
XX ID AAS59517 standard; DNA: 18796 BP.
XX AC AAS59517;
XX DT 13-FEB-2002 (first entry)
XX XX Propionibacterium acnes immunogenic protein encoding DNA #12.
XX DE
XX KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
XX KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
XX KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
XX KW dermatological; osteopathic; neuroprotectant; ds.
XX XX
XX OS Propionibacterium acnes.
XX PN WO200181581-A2.
XX PD 01-NOV-2001.
XX PF 20-APR-2001; 2001WO-US12865.
XX PR 21-APR-2000; 2000US-199047P.
XX PR 02-JUN-2000; 2000US-208841P.
XX PR 07-JUL-2000; 2000US-216747P.
XX PA (CORI-) CORIXA CORP.
XX PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
XX PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX WPI: 2001-616774/71.
XX PT Propionibacterium acnes polypeptides and nucleic acids useful for
XX PT vaccinating against and diagnosing infections, especially useful for
XX PT treating acne vulgaris -
XX PS Claim 1; SEQ ID No 12; 1069pp; English.
XX CC Sequences AAS59506-AAS59804 represent DNA molecules encoding
XX CC Propionibacterium acnes immunogenic polypeptides. The proteins and their
XX CC associated DNA sequences are used in the treatment, prevention and
XX CC diagnosis of medical conditions caused by P. acnes. The disorders include
XX CC SAPHO syndrome (synovitis, acne, pustulosis, hypertosis and
XX CC osteomyelitis), uveitis and endophthalmitis. P. acnes is also involved
XX CC in infections of bone, joints and the central nervous system, however it
XX CC is particularly involved in the inflammatory lesions associated with acne
XX CC vulgaris. A method for detecting the presence or absence of P. acnes in a
XX CC patient comprises contacting a sample with a binding agent that binds to
XX CC the proteins of the invention and determining the amount of bound protein
XX CC in the sample. The polypeptides may be used as antigens in the production
XX CC of antibodies specific for P. acnes proteins. These antibodies can be
XX CC used to downregulate expression and activity of P. acnes polypeptides and
XX CC therefore treat P. acnes infections. The antibodies may also be used as
XX CC diagnostic agents for determining P. acnes presence, for example, by
XX CC enzyme linked immunosorbent assay (ELISA). This sequence encodes the
XX CC polypeptides shown in AAU2344-42548 and AAU67479.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 18796 BP; 3824 A; 5690 C; 5570 G; 3711 T; 1 other;

Query Match 77.0%; Score 15.4; DB 23; Length 18796;
Best Local Similarity 94.1%; Pred. No. 4.4e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 cccgcccgaagccggg 20

||||| ||||| ||||| |||||

DB 9479 CCGGACGAGGCCGGG 9463

RESULT 14

AA136348/C

ID AA136348 standard; DNA; 19820 BP.

XX

AC AA136348;

XX

DT 08-JAN-2002 (first entry)

XX

DE Human musculoskeletal system related polynucleotide SEQ ID NO 2713.

XX

KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
antitumor; anticonvulsant; antibacterial; antifungal; antiparasitic;
cardiac; gene therapy; cancer; immune disorder; cardiovascular disorder;
neurological disease; infection; human; secreted protein;
musculoskeletal system; ds.

XX

OS Homo sapiens.

XX

PN WO200155367-A1.

XX

PD 02-AUG-2001.

XX

PF 17-JAN-2001; 2001WO-US01338.

XX

31-JAN-2000; 2000US-0179065.

PR

04-FEB-2000; 2000US-0180628.

PR

24-FEB-2000; 2000US-0184664.

PR

02-MAR-2000; 2000US-0186350.

PR

16-MAR-2000; 2000US-0189874.

PR

17-MAR-2000; 2000US-0190076.

PR

18-APR-2000; 2000US-0198123.

PR

19-MAY-2000; 2000US-0205515.

PR

07-JUN-2000; 2000US-0209467.

PR

28-JUN-2000; 2000US-0214886.

PR

30-JUN-2000; 2000US-0215135.

PR

07-JUL-2000; 2000US-0216647.

PR

07-JUL-2000; 2000US-0216880.

PR

11-JUL-2000; 2000US-0217487.

PR

11-JUL-2000; 2000US-0217496.

PR

14-JUL-2000; 2000US-0218290.

PR

26-JUL-2000; 2000US-0220963.

PR

26-JUL-2000; 2000US-0220964.

PR

14-AUG-2000; 2000US-0224518.

PR

14-AUG-2000; 2000US-0224519.

PR

14-AUG-2000; 2000US-0225213.

PR

14-AUG-2000; 2000US-0225214.

PR

14-AUG-2000; 2000US-0225266.

PR

14-AUG-2000; 2000US-0225267.

PR

14-AUG-2000; 2000US-0225268.

PR

14-AUG-2000; 2000US-0225270.

PR

14-AUG-2000; 2000US-0225447.

PR

14-AUG-2000; 2000US-0225757.

PR

14-AUG-2000; 2000US-0225758.

PR

14-AUG-2000; 2000US-0225759.

PR

18-AUG-2000; 2000US-0226279.

PR

22-AUG-2000; 2000US-0226681.

PR

22-AUG-2000; 2000US-0226688.

PR

22-AUG-2000; 2000US-0227182.

PR

PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241321.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.

PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA

XX Rosen CA, Barash SC, Ruben SM;
 XX WPI; 2001-451937/48.
 XX

XX Isolated polypeptide for treating, preventing and/ or prognosing
 PT disorders related to the musculoskeletal system including
 PT musculoskeletal cancers and also for testing and detection e.g.
 PT diagnosis -
 XX

PS Example 2; SEQ ID NO 2713; 781pp + Sequence Listing; English.
 XX

CC The invention relates to novel genes (AAL34669-AAL37666) and proteins
 CC (AB03087-AB04109) associated with the musculoskeletal system useful
 CC for preventing, treating or ameliorating medical conditions e.g. by
 CC protein or gene therapy. The genes are isolated from a range of human
 CC tissues disclosed in the specification. The nucleic acids, proteins,
 CC antibodies and (ant)agonists are useful in the diagnosis, treatment
 CC and prevention of: (a) cancer, e.g. breast and ovarian cancer and
 CC other cancers of the adrenal gland, bone, bone marrow, breast,
 CC gastrointestinal tract, liver, lung, or urogenital; (b) immune
 CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic
 CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
 CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis;
 CC (c) cardiovascular disorders such as myocardial ischaemias; (d) wound
 CC healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;
 CC and (f) infectious diseases such as viral, bacterial, fungal and
 CC parasitic infections.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX

XX Sequence 19820 BP; 6182 A; 3722 C; 4043 G; 5871 T; 2 other;

Query Match 77.0%; Score 15.4; DB 22; Length 19820;
 Best Local Similarity 94.1%; Pred. No. 4.4e+02;
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 cccggccgaagccggg 20
 |||||

Db 817 CCCGCCGAGGCGGG 801
 |||||

RESULT 15
 AAF30757/c
 ID AAF30757 standard; DNA; 47981 BP.
 XX
 AC AAF30757;
 XX
 DT 21-JUN-2001 (first entry)
 XX
 DE Micromonospora megalomicea megalomicin biosynthetic gene cluster.
 XX Micromonospora megalomicea subsp. nigra.
 KW Megalomicin; meg gene; polyketide synthase; antibiotic;
 KW motilide; antiparasitic; ds.
 XX
 OS Micromonospora megalomicea subsp. nigra.
 PH Key Location/Qualifiers
 FT CDS complement (1..144)
 FT /*tag= a
 FT /*partial
 FT /gene= "megBVI(megT)"
 FT /product= "TDP-4-keto-6-deoxyglucose-
 FT 2,3-dehydratase"
 FT /note= "encodes AAB82201"
 FT 928..2061
 FT /*tag= b
 FT /gene= "megDVI"
 FT /product= "TDP-4-keto-6-deoxyhexose 3,4-isomerase"
 FT /note= "encodes AAB82202"
 FT 2072..3382
 FT /*tag= c
 FT /gene= "megDI"
 FT /product= "TDP-megosamine glycosyltransferase"
 FT /note= "eryCIII homologue; encodes AAB802203"
 FT 3462..4634
 FT /*tag= d
 FT /gene= "megY"
 FT /product= "mycarose O-acyltransferase"
 FT /note= "encodes AAB82204"
 FT 4651..5775
 FT /*tag= e
 FT /gene= "megDII"
 FT /product= "TDP-3-keto-6-deoxyhexose
 FT 3-aminotransaminase"
 FT 5822..6595
 FT /note= "eryCI, DnrJ homologue, encodes AAB82205"
 FT /*tag= f
 FT /gene= "megDIII"
 FT /product= "daunosaminyl-N,N-dimethyltransferase"
 FT /note= "eryCVI homologue; encodes AAB82206"
 FT 6592..7197
 FT /*tag= g
 FT /gene= "megDIV"
 FT /product= "TDP-4-keto-6-deoxyhexose 3,5-epimerase"
 FT /note= "eryVIII, dnmU homologue, encodes AAB82207"
 FT 7220..8206
 FT /*tag= h
 FT /gene= "megDV"
 FT /product= "TDP-4-keto-6-deoxyhexose 4-ketoreductase"
 FT /note= "eryVIV, dnmV homologue; encodes AAB82208"
 FT complement (8228..9220)
 FT /*tag= i
 FT /gene= "megBII-1(megDVII)"
 FT /product= "TDP-4-keto-6-deoxyhexose 2,3-reductase"
 FT /note= "encodes AAB82209"
 FT complement (9226..10479)
 FT /*tag= j
 FT /gene= "megBY"
 FT /product= "TDP-mycarose glycosyltransferase"
 FT /note= "encodes AAB82210"
 FT complement (10483..11424)
 FT /*tag= k
 FT

```

FT FT /gene= "megBIV"
FT FT /product= "NDP-4-keto-6-deoxyhexose 4-ketoreductase"
FT FT /note= "encodes AAB82211"
FT FT /gene= "megA"
FT FT /tag= l
FT FT /product= "megalomicin 6-deoxyerythronolide B
FT FT     synthase 1"
FT FT /note= "polyketide synthase; encodes AAB82212"
FT FT 12505..13470
FT FT /tag= m
FT FT /gene= "megA"
FT FT /function= "AT-L"
FT FT 13576..13791
FT FT /tag= n
FT FT /gene= "megA"
FT FT /function= "ACP-L"
FT FT 13849..15126
FT FT /tag= o
FT FT /gene= "megA"
FT FT /function= "KS1"
FT FT 15427..16476
FT FT /tag= p
FT FT /gene= "megA"
FT FT /function= "AT1"
FT FT 17155..17694
FT FT /tag= q
FT FT /gene= "megA"
FT FT /function= "KR1"
FT FT 17947..18207
FT FT /tag= r
FT FT /gene= "megA"
FT FT /function= "ACPI"
FT FT 18268..19548
FT FT /tag= s
FT FT /gene= "megA"
FT FT /function= "KS2"
FT FT 19876..20910
FT FT /tag= t
FT FT /gene= "megA"
FT FT /function= "AT2"
FT FT 21517..22053
FT FT /tag= u
FT FT /gene= "megA"
FT FT /function= "KR2"
FT FT 22318..22575
FT FT /tag= v
FT FT /gene= "megA"
FT FT /function= "ACP2"
FT FT 22867..33555
FT FT /tag= w
FT FT /gene= "megAII"
FT FT /product= "megalomicin 6-deoxyerythronolide B
FT FT     synthase 2"
FT FT /note= "polyketide synthase, encodes AAB82213"
FT FT 22957..24237
FT FT /tag= x
FT FT /gene= "megAII"
FT FT /function= "KS3"
FT FT 24544..25581
FT FT /tag= y
FT FT /gene= "megAII"
FT FT /function= "AT3"
FT FT 26230..26733
FT FT /tag= z
FT FT /gene= "megAII"
FT FT /function= "KR3 (inactive)"
FT FT 26998..27258
FT FT /tag= aa
FT FT /gene= "megAII"
FT FT /function= "ACP3"
FT FT 27393..28590
FT FT /tag= ab

```

```

FT FT /gene= "megAII"
FT FT /function= "KS4"
FT FT 28897..29931
FT FT /tag= ac
FT FT /gene= "megAII"
FT FT /function= "AT4"
FT FT 29953..30477
FT FT /tag= ad
FT FT /gene= "megAII"
FT FT /function= "DH4"
FT FT 31396..32244
FT FT /tag= ae
FT FT /gene= "megAII"
FT FT /function= "ER4"
FT FT 32257..32799
FT FT /tag= af
FT FT /gene= "megAII"
FT FT /function= "KR4"
FT FT 33052..33312
FT FT /tag= ag
FT FT /gene= "megAII"
FT FT /function= "ACP4"
FT FT 33666..43271
FT FT /tag= ah
FT FT /gene= "megAIII"
FT FT /product= "megalomicin 6-deoxyerythronolide B
FT FT     synthase 3"
FT FT /note= "polyketide synthase; encodes AAB82214"
FT FT 22957..24237
FT FT /tag= ai
FT FT /gene= "megAIII"
FT FT /function= "KS5"
FT FT 24544..25581
FT FT /tag= aj
FT FT /gene= "megAIII"
FT FT /function= "AT5"
FT FT 26230..26733
FT FT /tag= ak
FT FT /gene= "megAIII"
FT FT /function= "KR5"
FT FT 26998..27258
FT FT /tag= al
FT FT /gene= "megAIII"
FT FT /function= "ACP5"
FT FT 27393..28590
FT FT /tag= am
FT FT /gene= "megAIII"
FT FT /function= "KS6"
FT FT 28897..29931
FT FT /tag= an
FT FT /gene= "megAIII"
FT FT /function= "AT6"
FT FT 29953..30477
FT FT /tag= ao
FT FT /gene= "megAIII"
FT FT /function= "KR6"
FT FT 31396..32244
FT FT /tag= ap

```

Query Match 77.0%; Score 15.4; DB 22; Length 47981;
 Best Local Similarity 94.1%; Pred. No. 4.1e+02;
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 cccggccgaagccggg 20
 |||||
 Db 6886 CCCGCCGAAGGTCGGG 6870

Search completed: June 19, 2002, 16:50:59
 Job time: 17517 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2002, 15:47:08 ; Search time 247.21 Seconds
(without alignments)
19.872 Million cell updates/sec

Title: US-09-462-955B-1_COPY_682_701

Perfect score: 20

Sequence: 1 ctaccggcgccgaagccggg 20

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA:*

1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*

2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*

3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*

4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*

5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*

6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	100.0	21	4	US-09-462-975-7
2	20	100.0	1291	4	US-09-462-975-1
3	15.8	79.0	3226	2	US-08-070-301-10
4	15.4	77.0	441529	4	US-09-103-840A-1
5	15.2	76.0	1520	1	US-08-225-477B-1
6	15.2	76.0	1520	5	PCT-US95-04353-1
7	15.2	76.0	50937	4	US-09-428-517-1
8	14.8	74.0	254	3	US-08-866-340-31
9	14.8	74.0	254	4	US-09-103-875-37
10	14.8	74.0	1002	1	US-08-457-707A-9
11	14.8	74.0	1002	1	US-08-812-025-9
12	14.8	74.0	1002	1	US-09-138-873A-9
13	14.8	74.0	3490	2	US-08-841-483-3
14	14.8	74.0	3490	4	US-09-382-911-3
15	14.8	74.0	4094	2	US-08-841-483-5
16	14.8	74.0	4094	4	US-09-382-911-5
17	14.8	74.0	4403765	4	US-09-103-840A-2
18	14.4	72.0	485	2	US-08-403-852D-15
19	14.4	72.0	485	3	US-08-510-646B-15
20	14.4	72.0	485	4	US-09-231-818-15
21	14.4	72.0	752	1	US-08-725-182C-1
22	14.4	72.0	752	3	US-09-013-150-1
23	14.4	72.0	752	4	US-09-359-503-1
24	14.4	72.0	752	4	US-09-062-422-1
25	14.4	72.0	752	4	US-09-396-184-1
26	14.4	72.0	752	4	US-08-937-263B-1
27	14.4	72.0	755	1	US-08-791-495-8

c	28	14.4	72.0	1879	6	5212296-5	Patent No. 5212296
	29	14.2	71.0	33	2	US-08-365-486A-2	Sequence 2, Appli
	30	14.2	71.0	33	4	US-08-880-342-2	Sequence 2, Appli
c	31	14.2	71.0	696	4	US-09-220-528-1	Sequence 1, Appli
	32	14.2	71.0	696	4	US-09-220-528-2	Sequence 2, Appli
c	33	14.2	71.0	1652	4	US-09-220-528-68	Sequence 68, Appli
	34	14.2	71.0	1652	4	US-09-220-528-68	Sequence 69, Appli
	35	14	70.0	3744	2	US-08-348-353-16	Sequence 16, Appli
	36	14	70.0	3744	2	US-08-465-965-16	Sequence 16, Appli
	37	14	70.0	3744	3	US-08-465-966-16	Sequence 16, Appli
c	38	13.8	69.0	23	1	US-08-308-949A-12	Sequence 12, Appli
	39	13.8	69.0	23	4	US-09-462-975-6	Sequence 6, Appli
c	40	13.8	69.0	212	1	US-08-435-684A-32	Sequence 32, Appli
	41	13.8	69.0	212	2	US-08-934-877A-32	Sequence 32, Appli
c	42	13.8	69.0	212	4	US-08-871-678C-32	Sequence 32, Appli
	43	13.8	69.0	423	1	US-08-470-179-171	Sequence 171, App
c	44	13.8	69.0	595	1	US-08-580-038-25	Sequence 25, Appli
	45	13.8	69.0	774	2	US-08-471-371-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1

US-09-462-975-7

; Sequence 7, Application US/09462975

; Patent No. 6303345

; GENERAL INFORMATION:

; APPLICANT: Rohde, Wolfgang

; APPLICANT: Becker, Dieter

; APPLICANT: Randles, John W.

; APPLICANT: Hehn, Alain

; APPLICANT: Salamini, Francesco

; TITLE OF INVENTION: USE OF VIRUS DNA AS A PROMOTER

; FILE REFERENCE: 23232.000301

; CURRENT APPLICATION NUMBER: US/09/462,975

; PRIOR FILING DATE: 2000-05-17

; PRIOR APPLICATION NUMBER: PCT/EP98/04345

; PRIOR FILING DATE: 1998-07-13

; PRIOR APPLICATION NUMBER: 19730502.4

; NUMBER OF SEQ ID NOS: 7

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 7

; LENGTH: 21

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: /note-synthetic

; OTHER INFORMATION: construct

US-09-462-975-7

Query Match 100.0%; Score 20; DB 4; Length 21;

Best Local Similarity 100.0%; Pred. No. 1.5;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctaccggcgccgaagccggg 20

Db 1 ctaccggcgccgaagccggg 20

RESULT 2

US-09-462-975-1

; Sequence 1, Application US/09462975

; Patent No. 6303345

; GENERAL INFORMATION:

; APPLICANT: Rohde, Wolfgang

; APPLICANT: Becker, Dieter

; APPLICANT: Randles, John W.

; APPLICANT: Hehn, Alain

; APPLICANT: Salamini, Francesco

; TITLE OF INVENTION: USE OF VIRUS DNA AS A PROMOTER

```
; FILE REFERENCE: 23232.0003U1
; CURRENT APPLICATION NUMBER: US/09/462,975
; CURRENT FILING DATE: 2000-05-17
; PRIOR APPLICATION NUMBER: PCT/EP98/04345
; PRIOR FILING DATE: 1998-07-13
; PRIOR APPLICATION NUMBER: 19730502.4
; PRIOR FILING DATE: 1997-07-16
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1291
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:/note=synthetic
; OTHER INFORMATION: construct
US-09-462-975-1

Query Match          100.0%; Score 20; DB 4; Length 1291;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctaccgcccgaagccgg 20
   |||||
Db 1072 ctaccgcccgaagccgg 1091

RESULT 3
US-08-070-301-10/c
; Sequence 10, Application US/08070301
; Patent No. 5871995
; GENERAL INFORMATION:
; APPLICANT: IIDA, Toshio
; APPLICANT: KAMINUMA, Toshihiko
; APPLICANT: FUSE, Yuka
; APPLICANT: TAJIMA, Masahiro
; APPLICANT: YANAGI, Mitsuo
; APPLICANT: OKAMOTO, Hiroshi
; APPLICANT: KISHIMOTO, Jiro
; APPLICANT: IFUKU, Ohji
; APPLICANT: KATO, Ichiro
; TITLE OF INVENTION: ENZYME PARTICIPATING IN C-TERMINAL
; TITLE OF INVENTION: AMIDATION, AND METHOD OF PREPARING SAME AND USE THEREOF
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wegner, Cantor, Mueller & Player, P.C.
; STREET: 1233 20th Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20036-8218
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/070,301
; FILING DATE: 24-MAY-1991
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 1-209687
; FILING DATE: 15-AUG-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 1-181933
; FILING DATE: 31-OCT-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 2-76331
; FILING DATE: 26-MAR-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 2-106412
; FILING DATE: 24-APR-1990
```

```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 2-205475
; FILING DATE: 02-AUG-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Player, William E.
; REGISTRATION NUMBER: 31,409
; REFERENCE/DOCKET NUMBER: P-450-22830
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-040
; TELEFAX: (202) 835-0605
; TELEX: 440706
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3226 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Rat
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2..831
; US-08-070-301-10

Query Match          79.0%; Score 15.8; DB 2; Length 3226;
Best Local Similarity 89.5%; Pred. No. 64;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctaccgcccgaagccgg 19
   |||||
Db 2585 CTACCGGGCGAAGCTGG 2567

RESULT 4
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
; US-09-103-840A-1

Query Match          77.0%; Score 15.4; DB 4; Length 4411529;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 cccgcccgaagccgg 19
   |||||
Db 1558394 cccgcccgaagccgg 1558409

RESULT 5
US-08-225-477B-1
; Sequence 1, Application US/08225477B
; Patent No. 5635370
; GENERAL INFORMATION:
```



```

; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 37
; LENGTH: 254
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-103-875-37

Query Match      74.0%; Score 14.8; DB 4; Length 254;
Best Local Similarity 88.9%; Pred No. 2.1e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 accggccgaagccggg 20
   ||| ||| ||| ||| |||
Db 66 ACCAGGCCCAAGGCCGGG 49

RESULT 10
US-08-457-797A-9
; Sequence 9, Application US/08457797A
; Patent No. 5689045
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Transgenic pathogen-resistant organism
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brumbaugh, Graves, Donohue and Raymond
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/457,797A
; FILING DATE: June 1, 1995
; ATTORNEY/AGENT INFORMATION:
; REGISTRATION NUMBER: 18,839
; REFERENCE/DOCKET NUMBER: A29542 FWC - 37/31335
; TELEPHONE: (212) 408-2500
; TELEFAX: (212) 765-2519
; TELEX: 238555
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1002 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Hordeum vulgare
; STRAIN: L.
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: 1..63
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 64..861
; OTHER INFORMATION: /codon_start= 64
; OTHER INFORMATION: /function= "chitinase"
; OTHER INFORMATION: /product= "26 kD preprotein of chitinase G (ChiG)"
; OTHER INFORMATION: /note= "antifungal activity, especially on
; OTHER INFORMATION: Trichoderma reesei and Fusarium sporotrichoides as
; OTHER INFORMATION: well as Rhizoctonia solani and Botrytis cinerea."
; FEATURE:
; NAME/KEY: 3'UTR
; LOCATION: 862..1002

```

```

; US-08-866-340-31/c
; Sequence 31, Application US/08866340
; Patent No. 6020318
; GENERAL INFORMATION:
; APPLICANT: Szyf, Moshe
; APPLICANT: Bigey, Pascal
; APPLICANT: Ramchandani, Shyam
; TITLE OF INVENTION: DNA METHYLTRANSFERASE GENOMIC
; TITLE OF INVENTION: SEQUENCES AND ANTISENSE OLIGONUCLEOTIDES
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HALE AND DORR LLP
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: United States of America
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/866,340
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Keown, Wayne A.
; REGISTRATION NUMBER: 33,923
; REFERENCE/DOCKET NUMBER: 106.101.187
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 526-6000
; TELEFAX: (617) 526-5000
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 254 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-866-340-31

Query Match      74.0%; Score 14.8; DB 3; Length 254;
Best Local Similarity 88.9%; Pred. No. 2.1e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 accggccgaagccggg 20
   ||| ||| ||| ||| |||
Db 66 ACCAGGCCCAAGGCCGGG 49

RESULT 9
US-09-103-875-37/c
; Sequence 37, Application US/09103875A
; Patent No. 6221849
; GENERAL INFORMATION:
; APPLICANT: Szyf, Moshe
; APPLICANT: Bigey, Pascal
; APPLICANT: Ramchandani, Shyam
; TITLE OF INVENTION: DNA METHYLTRANSFERASE GENOMIC SEQUENCES AND ANTISENSE
; TITLE OF INVENTION: OLIGONUCLEOTIDES
; FILE REFERENCE: 106101.194
; CURRENT APPLICATION NUMBER: US/09/103,875A
; CURRENT FILING DATE: 1998-06-24
; EARLIER APPLICATION NUMBER: 60/069,865
; EARLIER FILING DATE: 1997-12-17
; EARLIER APPLICATION NUMBER: 08/866,340
; EARLIER FILING DATE: 1997-05-30
; NUMBER OF SEQ ID NOS: 138

```

```

; OTHER INFORMATION: /partial
; OTHER INFORMATION: /note= "11 nucleotides at 3' end not shown"
; FEATURE:
; NAME/KEY: polyA_signal
; LOCATION: 905..910
; OTHER INFORMATION: /note= "potential polyadenylation
; OTHER INFORMATION: signal"
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 64..294
; OTHER INFORMATION: /note= "probable signal peptide
; OTHER INFORMATION: sequence"
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 298..312
; OTHER INFORMATION: /note= "probable signal peptide
; OTHER INFORMATION: sequence"
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 349..378
; OTHER INFORMATION: /note= "probable signal peptide
; OTHER INFORMATION: sequence"
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 466..588
; OTHER INFORMATION: /note= "probable signal peptide
; OTHER INFORMATION: sequence"
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 607..861
; OTHER INFORMATION: /note= "probable signal peptide
; OTHER INFORMATION: sequence"
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 133..861
; OTHER INFORMATION: /note= "probable signal peptide
; OTHER INFORMATION: sequence"
; US-08-457-797A-9

```

```

Query Match 74.0%; Score 14.8; DB 1; Length 1002;
Best Local Similarity 88.9%; Pred. No. 1.9e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

Oy 3 accgcgcgaagccggg 20
    ||| ||||| |||||
Db 677 ACCGGCGCGAGCGCGG 694

```

```

RESULT 11
US-08-812-025-9
; Sequence 9, Application US/08812025
; Patent No. 5804184
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Transgenic pathogen-resistant organism
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brumbaugh, Graves, Donohue and Raymond
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (BPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/812,025
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Tenser, Arthur
; REGISTRATION NUMBER: 18,839

```

```

; NAME: Kole, Lisa
; REGISTRATION NUMBER: 35,225
; REFERENCE/DOCKET NUMBER: A29542 FWC - 37/31335
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 408-2500
; TELEFAX: (212) 765-2519
; TELEX: 238555
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1002 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Hordeum vulgare
; STRAIN: L.
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: 1..63
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 64..861
; OTHER INFORMATION: /codon_start= 64
; OTHER INFORMATION: /function= "chitinase"
; OTHER INFORMATION: /product= "26 kD preprotein of chitinase G (ChiG)"
; OTHER INFORMATION: /note= "antifungal activity, especially on
; OTHER INFORMATION: Trichoderma reesei and Fusarium sporotrichoides as
; OTHER INFORMATION: well as Rhizoctonia solani and Botrytis cinerea."
; FEATURE:
; NAME/KEY: 3'UTR
; LOCATION: 862..1002
; OTHER INFORMATION: /partial
; OTHER INFORMATION: /note= "11 nucleotides at 3' end not shown"
; FEATURE:
; NAME/KEY: polyA_signal
; LOCATION: 905..910
; OTHER INFORMATION: /note= "potential polyadenylation
; OTHER INFORMATION: signal"
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 64..294
; OTHER INFORMATION: /note= "probable signal peptide
; OTHER INFORMATION: sequence"
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 298..312
; OTHER INFORMATION: /note= "probable signal peptide
; OTHER INFORMATION: sequence"
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 349..378
; OTHER INFORMATION: /note= "probable signal peptide
; OTHER INFORMATION: sequence"
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 466..588
; OTHER INFORMATION: /note= "probable signal peptide
; OTHER INFORMATION: sequence"
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 607..861
; OTHER INFORMATION: /note= "probable signal peptide
; OTHER INFORMATION: sequence"
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 133..861
; OTHER INFORMATION: /note= "probable signal peptide
; OTHER INFORMATION: sequence"
; US-08-812-025-9

```

```

Query Match 74.0%; Score 14.8; DB 1; Length 1002;
Best Local Similarity 88.9%; Pred. No. 1.9e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

Thu Jun 20 06:56:49 2002

LOCATION: 64..294
OTHER INFORMATION: /note= "probable signal peptide"
OTHER INFORMATION: sequence"
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 298..312
OTHER INFORMATION: /note= "probable signal peptide"
OTHER INFORMATION: sequence"
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 349..378
OTHER INFORMATION: /note= "probable signal peptide"
OTHER INFORMATION: sequence"
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 466..588
OTHER INFORMATION: /note= "probable signal peptide"
OTHER INFORMATION: sequence"
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 607..861
OTHER INFORMATION: /note= "probable signal peptide"
OTHER INFORMATION: sequence"
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 133..861
US-09-138-873A-9

Query Match 74.0%; Score 14.8; DB 4; Length 1002;
Best Local Similarity 88.9%; Pred. No. 1.9e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 acccgccgaagccggg 20
DB 677 ACCGGCCGCGAGCCGG 694

RESULT 13

US-08-841-483-3/c
; Sequence 3, Application US/08841483B
; Patent No. 5976875
; GENERAL INFORMATION:
; APPLICANT: Prescott, Steven M.
; APPLICANT: Bunting, Michaeline
; APPLICANT: Tang, Wen
; APPLICANT: Topham, Matthew
; TITLE OF INVENTION: Diacylglycerol Kinase Isoforms Epsilon and Zeta and
; TITLE OF INVENTION: Methods of Use Thereof
; FILE REFERENCE: 2037.2.1a
; CURRENT APPLICATION NUMBER: US/08/841,483B
; CURRENT FILING DATE: 1997-04-22
; EARLIER APPLICATION NUMBER: 60/016,210
; EARLIER FILING DATE: 1996-04-22
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 3490
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (89)..(2875)
US-08-841-483-3

Query Match 74.0%; Score 14.8; DB 2; Length 3490;
Best Local Similarity 88.9%; Pred. No. 1.7e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 acccgccgaagccggg 20
DB 3255 ACCGGCCGCGAGCCAGG 3238

QY 3 acccgccgaagccggg 20
DB 677 ACCGGCCGCGAGCCGG 694

RESULT 12

US-09-138-873A-9
; Sequence 9, Application US/09138873A
; Patent No. 6271438
; GENERAL INFORMATION:
; APPLICANT: Transgenic pathogen-resistant organism
; TITLE OF INVENTION: Transgenic pathogen-resistant organism
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Baker & Botts
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/138,873A
; FILING DATE: August 24, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Tenser, Arthur
; REGISTRATION NUMBER: 18,839
; NAME: Kote, Lisa
; REGISTRATION NUMBER: 35,225
; REFERENCE/DOCKET NUMBER: A29542 FWC - 37/31335
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 408-2500
; TELEFAX: (212) 765-2519
; TELEX: 238555
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1002 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Hordeum vulgare
; STRAIN: L.
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: 1..63
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 64..861
; OTHER INFORMATION: /codon_start= 64
; OTHER INFORMATION: /function= "chitinase"
; OTHER INFORMATION: /product= "26 kD preprotein of chitinase G (ChicG)"
; OTHER INFORMATION: /note= "antifungal activity, especially on
; OTHER INFORMATION: Trichoderma reesei and Fusarium sporotrichoides as
; OTHER INFORMATION: well as Rhizoctonia solani and Botrytis cinerea."
; FEATURE:
; NAME/KEY: 3'UTR
; LOCATION: 862..1002
; OTHER INFORMATION: /partial
; OTHER INFORMATION: /note= "11 nucleotides at 3' end not shown"
; FEATURE:
; NAME/KEY: polyA_signal
; LOCATION: 905..910
; OTHER INFORMATION: /note= "potential polyadenylation
; OTHER INFORMATION: signal"
; FEATURE:
; NAME/KEY: sig_peptide

RESULT 14

US-09-382-911-3/C
 ; Sequence 3, Application US/09382911
 ; Patent No. 6221658
 ; GENERAL INFORMATION:
 ; APPLICANT: Prescott, Steven M.
 ; APPLICANT: Bunting, Michaeline
 ; APPLICANT: Tang, Wen
 ; APPLICANT: Topham, Matthew
 ; TITLE OF INVENTION: Diacylglycerol Kinase Isoforms Epsilon and Zeta and
 ; TITLE OF INVENTION: Methods of Use Thereof
 ; FILE REFERENCE: 2037.2.1a
 ; CURRENT APPLICATION NUMBER: US/09/382,911
 ; PRIOR FILING DATE: 1999-08-25
 ; PRIOR APPLICATION NUMBER: 08/841,483
 ; PRIOR FILING DATE: 1997-04-22
 ; PRIOR APPLICATION NUMBER: 60/016,210
 ; PRIOR FILING DATE: 1996-04-22
 ; NUMBER OF SEQ ID NOS: 33
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 3
 ; LENGTH: 3490
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (89)..(2875)
 US-09-382-911-3

Query Match 74.0%; Score 14.8; DB 4; Length 3490;
 Best Local Similarity 88.9%; Pred. No. 1.7e+02;
 Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 acccgccggaaggccggg 20
 ||||| ||||| |||||
 Db 3255 ACCCGCCGAAGACGAG 3238

RESULT 15

US-08-841-483-5/C
 ; Sequence 5, Application US/08841483B
 ; Patent No. 5976875
 ; GENERAL INFORMATION:
 ; APPLICANT: Prescott, Steven M.
 ; APPLICANT: Bunting, Michaeline
 ; APPLICANT: Tang, Wen
 ; APPLICANT: Topham, Matthew
 ; TITLE OF INVENTION: Diacylglycerol Kinase Isoforms Epsilon and Zeta and
 ; TITLE OF INVENTION: Methods of Use Thereof
 ; FILE REFERENCE: 2037.2.1a
 ; CURRENT APPLICATION NUMBER: US/08/841,483B
 ; CURRENT FILING DATE: 1997-04-22
 ; EARLIER APPLICATION NUMBER: 60/016,210
 ; EARLIER FILING DATE: 1996-04-22
 ; NUMBER OF SEQ ID NOS: 33
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 5
 ; LENGTH: 4094
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (126)..(3479)
 US-08-841-483-5

Query Match 74.0%; Score 14.8; DB 2; Length 4094;
 Best Local Similarity 88.9%; Pred. No. 1.7e+02;
 Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 acccgccggaaggccggg 20
 ||||| ||||| |||||
 Db 3859 ACCCGCCGAAGACGAG 3842
 Search completed: June 19, 2002, 15:49:00
 Job time: 13798 sec

us-09-462-955b-1_copy_682_701.rni

Thu Jun 20 06:56:49 2002

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2002, 14:04:08 ; Search time 7489.97 Seconds
(without alignments)
36.040 Million cell updates/sec

Title: US-09-462-955B-1_COPY_682_701

Perfect score: 20

Sequence: 1 ctaccgcggcgaagccggg 20

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	17	85.0	448	9	BE114197
2	17	85.0	491	9	BE114196
3	16.8	84.0	612	10	BE238940
4	16.4	82.0	786	10	BE679019
5	16.4	82.0	849	9	AL539606
6	16.4	82.0	996	10	BE260627
7	16.4	82.0	1172	10	BM467932
8	16	80.0	490	10	BG320293
9	15.8	79.0	309	10	BF286795
10	15.8	79.0	313	12	B18042
11	15.8	79.0	318	10	BF779841
12	15.8	79.0	319	10	BF779842
13	15.8	79.0	422	9	AW142153
14	15.8	79.0	426	10	BG298646
15	15.8	79.0	438	10	BM376491
16	15.8	79.0	470	12	B18122
17	15.8	79.0	473	10	BG417064

c	18	15.8	79.0	507	9	AW919607
19	15.8	79.0	552	10	BG846215	
20	15.8	79.0	575	10	BF103096	
c	21	15.8	79.0	669	12	AG093531
22	15.8	79.0	678	10	BG846216	
23	15.8	79.0	717	10	BE454137	
24	15.8	79.0	876	10	BG181627	
c	25	15.8	79.0	877	10	BI908420
26	15.8	79.0	909	12	AZ210881	
27	15.8	79.0	920	12	CNS02261	
28	15.8	79.0	980	10	BF780277	
29	15.8	79.0	1006	12	CNS01TTB	
30	15.8	79.0	1156	10	BE791915	
31	15.8	79.0	1214	12	AG086086	
32	15.4	77.0	147	9	AA283327	
33	15.4	77.0	232	9	BB571165	
c	34	15.4	77.0	352	9	AW337668
35	15.4	77.0	377	9	BB813476	
36	15.4	77.0	416	9	AA997260	
c	37	15.4	77.0	417	10	BM199463
38	15.4	77.0	438	9	AW619475	
39	15.4	77.0	478	9	AW916542	
c	40	15.4	77.0	483	9	BB697893
41	15.4	77.0	500	12	BH353262	
c	42	15.4	77.0	617	12	BH549318
c	43	15.4	77.0	640	10	BM439271
c	44	15.4	77.0	685	10	BI559944
45	15.4	77.0	717	9	AI963704	

ALIGNMENTS

RESULT 1

BE114197
LOCUS
DEFINITION
UI-R-CA0-axo-b-10-0-UI.s1 UI-R-CA0 Rattus norvegicus cDNA clone
UI-R-CA0-axo-b-10-0-UI 3', mRNA sequence.
BE114197
ACCESSION
VERSION
KEYWORDS
SOURCE
Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
AUTHORS
TITLE
Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL
MEDLINE
COMMENT
Genome Res. 6 (9), 791-806 (1996)
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
Oligo-dt track not found, Not I site shown in beginning of sequence
is likely internal to the message. cDNA Library Preparation: M.B.
Soares Lab Clone distribution: Clones will be available through
Research Genetics (www.resgen.com)
Seq primer: M13 forward
POLYA=NO.
Location/Qualifiers
1. .448
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-CA0-axo-b-10-0-UI"
/clone_lib="UI-R-CA0"
/lab_host="DH10B (Life Technologies)"

AW919607 EST350911
BG846215 1024012G0
BF103096 601648433
AG093531 Pan trogl
BG846216 1024012G0
BE454137 HVSMEH008
BG181627 RST477 AT
BF098420 603067366
AZ210881 SP_0153_B
AL220563 Tetraodon
BF780277 602103378
AL166952 Tetraodon
BE791915 601585811
AG086086 Pan trogl
AA283327 RTH084 HT
BB571165 BB571165
AW337668 xv66e05.x
BB813476 BB813476
AA997260 UI-R-CO-h
BM199463 C0459H04-
AW619475 7534 MARC
AW916542 EST347846
BB697893 BB697893
BH353262 CH230-68L
BH549318 BGN0677R
BM439271 GIT000007
BI559944 603253407
AI963704 wr64h07.x

Thu Jun 20 06:56:49 2002

us-09-462-955b-1_copy_682_701.rst

/note="Vector: p7T3D-Pac (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; The UI-R-CAO library is a subtracted library derived from the following tissues: thalamus, cerebellum, hypothalamus, medulla, pons, midbrain, cerebral cortex, corpus striatum, testis, and hippocampus. For a detailed description of the library from which this clone was derived, please visit our web site at ratest.eng.uiowa.edu. The subtraction has been previously described in (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)

Genome Research 6:791-806, 1996)

BASE COUNT 69 a 125 c 171 g 81 t 2 others

ORIGIN

Query Match 85.0%; Score 17; DB 9; Length 448;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 cccgcccgaagccggg 20
|||||

Db 32 CCCGCCGAGGCCGGG 48

RESULT 2
LOCUS BE114196 491 bp mRNA linear EST 13-JUN-2000
DEFINITION UI-R-CAO-axo-b-09-0-UI.s1 UI-R-CAO Rattus norvegicus cDNA clone

ACCESSION BE114196
VERSION BE114196
KEYWORDS EST.
SOURCE BE114196.1 GI:8506301

ORGANISM Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 491)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: mscares@blue.weeg.uiowa.edu
Oligo-dt track not found, Not I site shown in beginning of sequence is likely internal to the message. cDNA Library Preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA-No.

FEATURES
source
1..491
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-CAO-axo-b-09-0-UI"
/clone_lib="UI-R-CAO"
/lab_host="DH10B (Life Technologies)"
/note="Vector: p7T3D-Pac (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; The UI-R-CAO library is a subtracted library derived from the following tissues: thalamus, cerebellum, hypothalamus, medulla, pons, midbrain, cerebral cortex, corpus striatum, testis, and hippocampus. For a detailed description of the library from which this clone was derived, please visit our web site at ratest.eng.uiowa.edu. The subtraction has been previously described in (Bonaldo, Lennon and Soares,

Genome Research 6:791-806, 1996)

BASE COUNT 77 a 142 c 179 g 92 t 1 others

ORIGIN

Query Match 85.0%; Score 17; DB 9; Length 491;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 cccgcccgaagccggg 20
|||||

Db 32 CCCGCCGAGGCCGGG 48

RESULT 3
LOCUS BE238940/c 612 bp mRNA linear EST 11-JUL-2000
DEFINITION MD0596 Meloidogyne incognita J2 (#MD99-1) Meloidogyne incognita cDNA clone 3438 5' similar to SEC-2 protein (Y09293), mRNA sequence.

ACCESSION BE238940
VERSION BE238940.1 GI:9033904
KEYWORDS EST.
SOURCE southern root-knot nematode.

ORGANISM Meloidogyne incognita
Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
Tylenchoidea; Heteroderidae; Meloidogyninae; Meloidogyne.

REFERENCE 1 (bases 1 to 612)
AUTHORS Dautova,M., Gommers,F.J., Bakker,J. and Smant,G.
TITLE 5' end expressed sequence tags from Meloidogyne incognita parasitic J2 cDNA library

JOURNAL Unpublished (2000)
COMMENT Contact: Smant G / Dautova M

LABORATORY OF Nematology
Wageningen University and Research Center
Box 8123, Binnenhaven 10, 6709 PD Wageningen, The Netherlands
Tel: 31 317 485 254
Fax: 31 317 484 254
Email: Geert.Smant@medew.nema.wau.nl,
Makedonka.Dautova@medew.nema.wau.nl
Insert Length: 612 Std Error: 0.00
Seq primer: T7 promoter primer
High quality sequence stop: 612.

FEATURES
source

1..612
/organism="Meloidogyne incognita"
/db_xref="taxon:6306"
/clone="3438"
/clone_lib="Meloidogyne incognita J2 (#MD99-1)"
/dev_stage="second stage parasitic juveniles (J2)"
/note="Vector: pMAK1; Site.1: Sfi IA; Site.2: Sfi IB; cDNA was synthesized using SMART III oligo, CDS III oligo d(T)30 (Clontech) and Superscript II reverse transcriptase (Life Technology). cDNA clones were size fractionated and directionally ligated in the Sfi IA restriction site at 5' end and Sfi IB at 3' end of pMAK1. pMAK1 was derived from the plasmid pcDNA II (Invitrogen)."

BASE COUNT 219 a 146 c 120 g 127 t

ORIGIN

Query Match 84.0%; Score 16.8; DB 10; Length 612;
Best Local Similarity 90.0%; Pred. No. 1.6e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctaccggccgaagccggg 20
|||||

Db 470 CTACGGCGCGAGGCCGGG 451

RESULT 4
BF679019

LOCUS
 DEFINITION 602153663F1 NIH_MGC_83 Homo sapiens cDNA clone IMAGE:4294839 5',
 mRNA sequence.
 ACCESSION BF679019
 VERSION BF679019.1 GI:11952914
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 786)
 NIH-MGC http://mgc.nci.nih.gov/.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 CONTACT: Robert Strausberg, Ph.D.
 EMAIL: cgapbs-r@mail.nih.gov
 Tissue Procurement: CLONETECH Laboratories, Inc.
 cDNA Library Preparation: CLONETECH Laboratories, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 Cloning by: Incyte Genomics, Inc.
 Cloning Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLCM144 row: m column: 16
 High quality sequence stop: 564.

FEATURES
 source
 1..786
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4294839"
 /lab_host="NIH_MGC_83"
 /note="Organ: prostate; Vector: pDNR-LIB (Clontech);
 Site 1: SfiI (ggcgctcgcc); Site 2: SfiI (ggccattatggcc
); 5' and 3' adaptors were used in cloning as follows: 5'
 adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor
 (where B = A, C, or G and N = A, C, G, or T). Average
 insert size 1.4 kb (range 0.5-4.0 kb). 14/15 colonies
 contained inserts by PCR. This library was enriched for
 full-length clones and was constructed by Clontech
 Laboratories (Palo Alto, CA)."
 218 a 174 c 167 g 227 t

BASE COUNT
 ORIGIN

Query Match 82.0%; Score 16.4; DB 10; Length 786;
 Best Local Similarity 94.4%; Pred. No. 2.4e+03;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 3 acccgccgaagccggg 20
 |||||
 Db 703 ACCCGCCGAGCGGGG 720

RESULT 5
 ACCESSION AL539606
 LOCUS AL539606 LTI_FL013_FBrnl Homo sapiens cDNA clone CSODF036YD18 5
 DEFINITION prime, mRNA sequence.
 ACCESSION AL539606
 VERSION AL539606.1 GI:12868977
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 849)
 Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
 Full-length cDNA libraries and normalization
 Unpublished (2001)
 CONTACT: Genoscope
 Genoscope - Centre National de Sequencage

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

BP 191 91006 EVRY cedex - France
 Email: seget@genoscope.cns.fr, Web : www.genoscope.cns.fr.
 FEATURES
 source
 1..849
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="CSODF036YD18"
 /dev_stage="pooled tissue from post conception fetuses (20
 week, 24 week and 26 week)"
 /lab_host="DH10B"
 /note="Organ: Fetal brain; Vector: pCMVSPORT 6; 1st strand
 cDNA was primed with a NotI-oligo(dT) primer. Five prime
 end enriched, double-stranded cDNA was digested with Not I
 and cloned into the Not I and Eco RV sites of the
 pCMVSPORT 6 vector. Library was constructed by Life
 Technologies. Contact : Feng Liang Life Technologies, a
 division of Invitrogen 9800 Medical Center Drive Rockville
 , Maryland 20850, USA Fax : (1) 301 610 8371 Email :
 fliang@lifetech.com URL :
 http://fulllength.invitrogen.com"
 BASE COUNT 189 a 187 c 267 g 205 t 1 others
 ORIGIN

Query Match 82.0%; Score 16.4; DB 9; Length 849;
 Best Local Similarity 85.0%; Pred. No. 2.4e+03;
 Matches 17; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 1 ctaccggccgaagccggg 20
 |||||
 Db 412 CGACGCGCCGAGCGGGG 431

RESULT 6
 ACCESSION BE260627/c
 LOCUS BE260627 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3161405 5',
 DEFINITION mRNA sequence.
 ACCESSION BE260627
 VERSION BE260627.1 GI:9132066
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 996)
 NIH-MGC http://mgc.nci.nih.gov/.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 CONTACT: Robert Strausberg, Ph.D.
 EMAIL: cgapbs-r@mail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Ling Hong/Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
 Plate: LLCM119 row: c column: 06
 High quality sequence start: 4
 High quality sequence stop: 661.
 Location/Qualifiers
 1..996
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3161405"
 /clone_lib="NIH_MGC_19"
 /tissue_type="neuroblastoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: brain; Vector: pOTB7; Site 1: XhoI; Site 2:
 EcoRI; cDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGCACGAG(G). Library constructed by Ling Hong

FEATURES
 source
 1..996
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3161405"
 /clone_lib="NIH_MGC_19"
 /tissue_type="neuroblastoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: brain; Vector: pOTB7; Site 1: XhoI; Site 2:
 EcoRI; cDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGCACGAG(G). Library constructed by Ling Hong

in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

BASE COUNT 246 a 271 c 263 g 216 t
ORIGIN

Query Match 82.0%; Score 16.4; DB 10; Length 996;
Best Local Similarity 94.4%; Pred. No. 2.4e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 taccggcggaagccgg 19
||||| ||||| ||||| ||||| |||||

Db 760 TACCCGCCAAGGCCGG 743

RESULT 7
BM467932 1172 bp mRNA linear EST 05-FEB-2002
LOCUS
DEFINITION AGENCOURT_6437854 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5532832
5', mRNA sequence.

ACCESSION BM467932
VERSION
KEYWORDS
SOURCE EST.
ORGANISM human.

REFERENCE Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1172)
TITLE NTH-MGC http://mgc.ncl.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov

Tissue Procurement: AFCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12216 row: 1 column: 17
High quality sequence start: 5
High quality sequence stop: 638.

FEATURES
Location/Qualifiers
1..1172

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5532832"
/clone_lib="NIH_MGC_71"
/tissue_type="leiomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site:1: NotI;
Site:2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 2.1 kb."

BASE COUNT 268 a 384 c 299 g 221 t
ORIGIN

Query Match 82.0%; Score 16.4; DB 10; Length 1172;
Best Local Similarity 94.4%; Pred. No. 2.4e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 accggcggaagccgg 20
||||| ||||| ||||| ||||| |||||

Db 940 ACTCGCCGAAGGCCGG 957

RESULT 8
BG320293 490 bp mRNA linear EST 27-FEB-2001
LOCUS
DEFINITION Zm03_12e11_A Zm03_AAFc_ECORC_cold_stressed_maize_seedlings Zea mays

cdna clone Zm03_12e11, mRNA sequence.

ACCESSION BG320293
VERSION BG320293.1 GI:13149971

KEYWORDS EST.
SOURCE Zea mays.

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.

REFERENCE 1 (bases 1 to 490)
AUTHORS Simmonds,J.A., Singh,J.A., Piche,C., Cass,L., Couroux,P., De Moors,A., Harris,L.J., Hattori,J.I., Ouellet,T., Robert,L.S., Sprott,D. and Tinker,N.A.

TITLE Expressed Sequence Tags from Cold-Stressed Maize Seedlings Grown Under High Light Intensity

JOURNAL Unpublished (2001)

COMMENT

Contact: Singh,J.A.

Eastern Cereal and Oilseed Research Centre

Agriculture and Agri-food Canada

960 Carling Avenue, Bldg. 20, Ottawa, Ontario, K1A 0C6, Canada

Tel: (613) 759-1662

Fax: (613) 759-1701

Email: singhja@em.agr.ca.

FEATURES

source

Location/Qualifiers
1..490

/organism="Zea mays"

/cultivar="CO328"

/db_xref="taxon:4577"

/clone="Zm03_12e11"

/clone_lib="Zm03_AAFc_ECORC_cold_stressed_maize_seedlings"

/tissue_type="Leaf, crown"

/dev_stage="4-leaf"

/note="Vector: Bluescript SK+/XhoI-EcoRI; Site:1: Eco RI;
Site:2: Xho I; Corn seedlings at 4-leaf stage were exposed
to low temperature/high light (100C/700-800uE/mz/s) for 4
days. Plants were grown/treated by J. Simmonds/L. Cass.
Library prepared by C. Piche using Stratagene kit."

BASE COUNT 56 a 168 c 171 g 71 t

ORIGIN

Query Match 80.0%; Score 16; DB 10; Length 490;
Best Local Similarity 80.0%; Pred. No. 3.4e+03;
Matches 16; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctaccggcggaagccgg 20
||||| ||||| ||||| ||||| |||||

Db 135 CCACCCGCCGATGCCGG 116

RESULT 9

BF286795/c

LOCUS

DEFINITION BF286795 Rat Gene Index, normalized rat, Rattus norvegicus CDNA

ACCESSION BF286795

VERSION BF286795.1 GI:11217865

KEYWORDS EST.

SOURCE Norway rat.

ORGANISM

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

REFERENCE 1 (bases 1 to 309)

AUTHORS Malek,R.L., Cho,J., Lee,Y., Karamycheva,S., Parvizi,B., Perle,G.,

Sultana,R., Tsai,J., White,J., Quackenbush,J. and Lee,N.H.

TITLE Generation of ESTs from Normalized Rat Embryo, Bonto Soares

JOURNAL Unpublished (2000)

COMMENT Other ESTs: EST451385

Contact: Lee, NH

The Institute for Genomic Research

9712, Medical Center Drive, Rockville, MD 20850, USA

Tel: (301)-838-3529

Query Match 79.0%; Score 15.8; DB 12; Length 313;

RESULT 12

1779842	OCUS	319 bp	linear	EST 26-SEP-2001
1779842	DEFINITION	EBem07.SQ001.C22.R IGF barley EBem07 library	Hordeum vulgare	cdna
	DESCRIPTION	clone EBem07.SQ001.C22.5, mRNA sequence.		
	ACCESSION	1779842		
	VERSION	1779842.1		
	KEYWORDS	GI:15782734		
	KEYWORDS	EST.		

SOURCE
ORGANISM
barley.
Hordeum vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticeae; Hordeum.
1 (bases 1 to 319)
Hedley, P., Liu, H., Caldwell, D., McCallum, N., Mudie, S., Cardle, L.,
Ramsay, L., Machray, G., Marshall, D.F.M. and Waugh, R.
Development of Barley Transcriptome Resources
Unpublished (2001)
Contact: Waugh R
Unit of Genomics
Scottish Crop Research Institute
Invergowrie, Dundee, DD2 5DA, Scotland, UK
Tel: 00 44 1382 562731
Fax: 00 44 1382 562426
Email: rwaugh@scri.sari.ac.uk
All sequence has a Phred quality score of 20 or over
Seq primer: M13 reverse
Location/Qualifiers
1. 319
/organism="Hordeum vulgare"
/cultivar="Optic"
/db_xref="taxon:4513"
/clone="EBem07_S0001_C22"
/clone_lib="IGF Barley EBem07 library"
/tissue_type="Embryo"
/dev_stage="28 days post anthesis"
/lab_host="DH10B"
/note="Vector: pSPORT1; Site_1: Sal I; Site_2: Not I;
Non-normalised library, directionally cloned into pSPORT1.
Derived from embryos dissected from developing grains (28
days post anthesis) in glasshouse grown barley plants.
Developed as part of the barley transcriptome resources of
BSRC/SEERAD funded cereal IGF (Investigating Gene
Function) project."
BASE COUNT 61 a 91 c 86 g 81 t
ORIGIN

Query Match 79.08; Score 15.8; DB 10; Length 319;
Best Local Similarity 89.5%; Pred. No. 4.1e+03;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctaccgcgcgaagccgg 19
||||| |||||
Db 34 CTACCCGCGGAATGCCGG 52

RESULT 13
AW142153/c
LOCUS
DEFINITION
AW142153 422 bp mRNA linear EST 30-OCT-1999
EST292377 Normalized rat heart, Bento Soares Rattus sp. cDNA clone
GRIAG83 5' end similar to peptidylglycine alpha-amidating
monooxygenase, mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
Rattus sp.
Rattus sp.
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 422)
Lee, N.H., Glodek, A., Chandra, I., Mason, T.M., Quackenbush, J.,
Kerlavage, A.R. and Adams, M.D.
Rat Genome Project: Generation of a Rat EST (EST) Catalog & Rat
Gene Index
Unpublished (1998)
Other ESTs: EST292376 TC89750
Contact: Lee, NH
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA

tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
For clone availability, additional sequence and expression
information related to this EST please check the TIGR Rat Gene
Index (<http://www.tigr.org/tadb/rji/rji.html>). To order a clone
contact the ATCC (<http://www.atcc.org/atcc.html>).
Seq primer: M13 Reverse.
Location/Qualifiers
1. 422
/organism="Rattus sp."
/db_xref="taxon:10118"
/clone="GRIAG83"
/clone_lib="Normalized rat heart, Bento Soares"
/note="Organ: heart; Vector: pT7T3pac; Site_1: EcoRI;
Site_2: NotI"
BASE COUNT 85 a 118 c 98 g 121 t
ORIGIN

Query Match 79.08; Score 15.8; DB 9; Length 422;
Best Local Similarity 89.5%; Pred. No. 4.2e+03;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctaccgcgcgaagccgg 19
||||| |||||
Db 83 CTACCCGCGCAAGGCTGG 65

RESULT 14
BG298646/c
LOCUS
DEFINITION
BG298646 426 bp mRNA linear EST 21-FEB-2001
602396847F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:4511794 5',
mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
BG298646
BG298646.1 GI:13063508
EST.
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 426)
NTH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: The Cepko Laboratory
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM10395 row: m column: 11
High quality sequence stop: 327.
Location/Qualifiers
1. 426
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:4511794"
/clone_lib="NIH_MGC_94"
/tissue_type="retina"
/lab_host="DH10B (phage-resistant)"
/note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 3.3 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 89 a 127 c 129 g 81 t
ORIGIN

Query Match 79.0%; Score 15.8; DB 10; Length 426;
 Best Local Similarity 89.5%; Pred. No. 4.2e-03;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctaccggcggaaggccgg 19
 ||||| | | | | | | | |
 Db 359 CTACCGCGGCAAGGCGG 341

RESULT 15

BM376491
 LOCUS
 DEFINITION EBem05_SQ002_G02_R IGF Barley EBem05 library linear EST 10-JAN-2002
 clone EBem05_SQ002_G02 5', mRNA sequence.

ACCESSION BM376491
 VERSION
 KEYWORDS EST.
 SOURCE BM376491.1 GI:18119881

SOURCE

ORGANISM barley.
 Hordeum vulgare
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
 ; Triticeae; Hordeum.

REFERENCE

AUTHORS Hedley, P., Liu, H., Caldwell, D., McCallum, N., Mudie, S., Cardle, L.,
 Ramsay, L., Machray, G., Marshall, D.F.M. and Waugh, R.

TITLE

JOURNAL Development of Barley Transcriptome Resources

COMMENT

Unpublished (2001)
 Contact: Waugh R
 Unit of Genomics
 Scottish Crop Research Institute
 Invergowrie, Dundee, DD2 5DA, Scotland, UK
 Tel: 00 44 1382 562731
 Fax: 00 44 1382 562426
 Email: rwaugh@scri.sari.ac.uk
 All sequence has a Phred quality score of 20 or over
 Seq primer: M13 reverse.

FEATURES

Location/Qualifiers

1..438
 /organism="Hordeum vulgare"
 /cultivar="Optic"
 /db_xref="taxon:4513"
 /clone="EBem05_SQ002_G02"
 /clone_lib="IGF Barley EBem05 library"
 /tissue_type="Embryo"
 /dev_stage="14 days post anthesis"
 /lab_host="DH10B"
 /note="Vector: pSPORT1; Site_1: Sal I; Site_2: Not I;
 Non-normalised library, directionally cloned into pSPORT1.
 Derived from embryos dissected from developing grains (14
 days post anthesis) in glasshouse grown barley plants.
 Developed as part of the barley transcriptome resources of
 BBSRC/SEERAD funded cereal IGF (Investigating Gene
 Function) project."

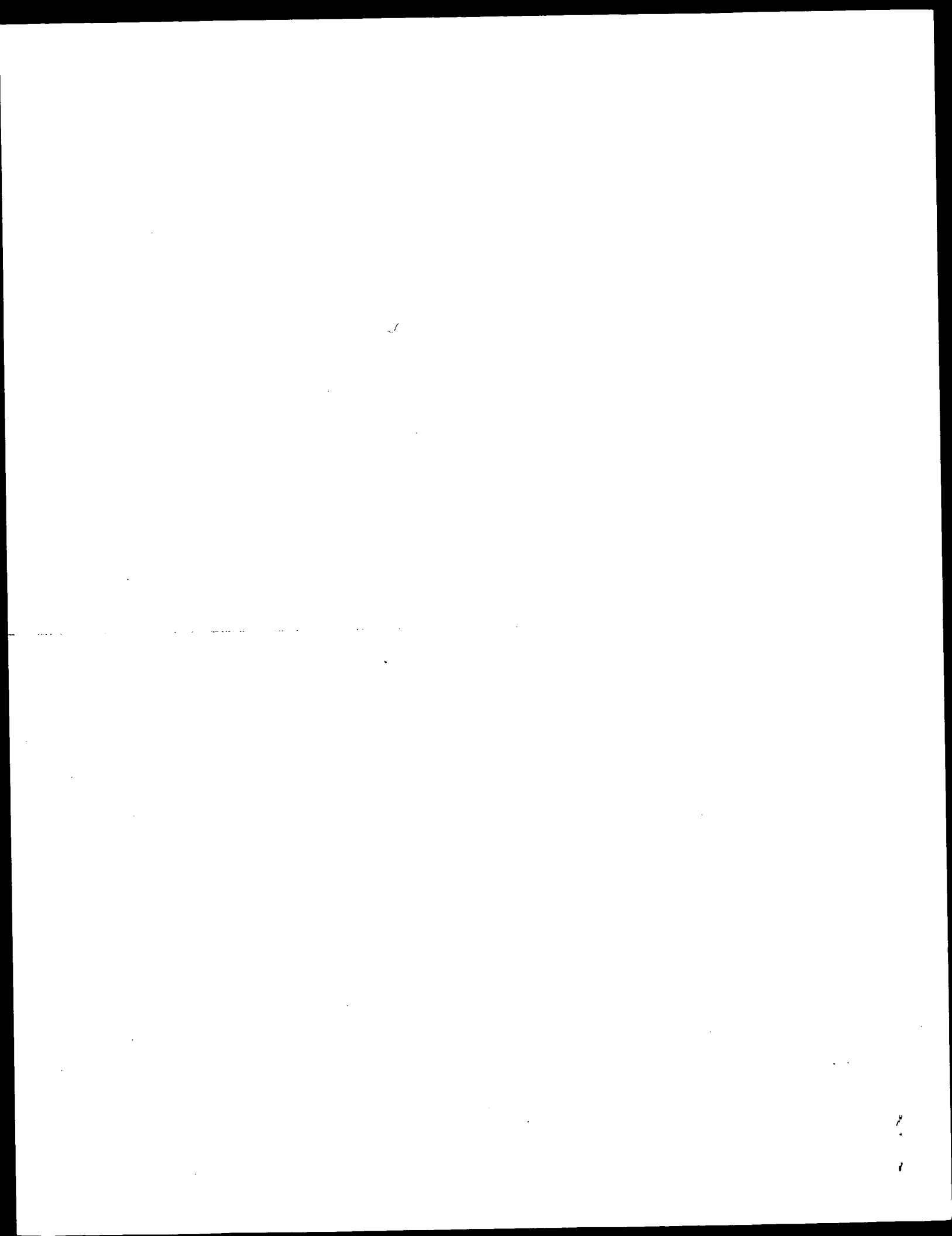
BASE COUNT 87 a 170 c 138 g 43 t
 ORIGIN

Query Match

Best Local Similarity 79.0%; Score 15.8; DB 10; Length 438;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ctaccggcggaaggccgg 19
 ||||| | | | | | | | |
 Db 361 CTACCGCGGCAAGGCGG 379

Search completed: June 19, 2002, 14:04:12
 Job time: 7510 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2002, 15:43:19 ; Search time 5904.86 Seconds
(without alignments)
2767.826 Million cell updates/sec

Title: US-09-462-955B-1_COPY_211_991

Perfect score: 781

Sequence: 1 atcgagagccagcgagcga.....ccccaacctctgtaacccc 781

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_hgt.*

3: gb_in.*

4: gb_em.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vt.*

15: gb_vt.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vi.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
------------	-------	-------------	--------	-------	-------------

RESULT 1

ARI12770	ARI12770	Sequence 1	from patent	US 6303345.
LOCUS	ARI12770	Sequence 1	from patent	US 6303345.
DEFINITION	ARI12770	Sequence 1	from patent	US 6303345.
ACCESSION	ARI12770	Sequence 1	from patent	US 6303345.
VERSION	ARI12770.1	GT:17912261		
KEYWORDS	Unknown.			
SOURCE	Unknown.			
ORGANISM	Unknown.			
REFERENCE	1 (bases 1 to 1291)			
AUTHORS	Rohde,W., Becker,D., Randles,J.W., Hehn,A. and Salamini,F.			
TITLE	Use of a virus DNA as promoter			
JOURNAL	Patent: US 6303345-A 1 16-OCT-2001;			
FEATURES	Location/Qualifiers			
source	1..1291			
BASE COUNT	336 a	323 c	332 g	300 t
ORIGIN				

ALIGNMENTS

1	691	88.5	1291	6	ARI12770	ARI12770 Sequence
2	691	88.5	1291	14	CFDCG	M29963 Coconut fol
3	116.4	14.9	1007	14	NYV5966	AJ132187 faba bean
4	114.8	14.7	1004	14	NYV5966	AJ005966 faba bean
5	114.2	14.6	1096	6	AR063452	Sequence
6	114.2	14.6	1096	14	BYTV2	L32167 Banana bunc
7	112.6	14.4	1091	6	AR063453	Sequence
8	112.4	14.4	1095	14	AF216222	ARI12770
9	99	12.7	1111	14	BBU12586	AF216222 Banana bu
10	98.6	12.6	1127	14	BBU12587	U12586 Banana bunc
11	95	12.2	1000	14	AB000922	U12587 Banana bunc
12	95	12.2	1002	14	FBNYREP	AB000922 Milk vetc
13	95	12.2	1109	14	AF216221	X80879 FBMYV gene
14	93	11.9	1004	14	NYV5968	AF216221 Banana bu
15	92.8	11.9	1007	14	AB000920	AJ005968 faba bean
16	92.8	11.9	1015	14	FBE132185	AB000920 Milk vetc
17	92.6	11.9	1106	6	AR063451	AJ132185 faba bean
18	92.6	11.9	1106	14	BYTV1	AR063451 Sequence
19	92.2	11.8	1009	14	AB000921	L32166 Banana bunc
20	90.8	11.6	1095	14	U02312	AB000921 Milk vetc
21	90.4	11.6	1014	14	NYV5964	U02312 Banana bunc
22	89.4	11.4	1375	14	CLE132345	AJ005964 faba bean
23	89	11.4	1017	6	AR145398	AJ132345 Nanovirus
24	89	11.4	1017	14	SCU16735	ARI145398 Sequence
25	88.4	11.3	1022	6	ARI145394	U16735 Subterranea
26	88.4	11.3	1022	14	SCU16731	ARI45394 Sequence
27	88	11.3	1022	14	AB009047	U16731 Subterranea
28	87	11.1	1367	14	AYE238493	AB009047 Milk vetc
29	85.4	10.9	1376	14	CLE132344	AJ238493 Nanovirus
30	62.8	8.0	1110	6	AR010230	AJ132344 Nanovirus
31	59.6	7.6	982	6	ARI12007	AR010230 Sequence
32	59.6	7.6	1110	6	ARI12027	ARI12007 Sequence
33	59.6	7.6	1110	14	BBU18077	ARI12027 Sequence
34	59.6	7.6	1111	6	AR010225	U18077 Banana bunc
35	59.6	7.6	1111	6	AR010228	AR010225 Sequence
36	59.6	7.6	1111	6	AR010240	AR010228 Sequence
37	59.6	7.6	1111	14	AF102780	AR010240 Sequence
38	59.6	7.6	1111	14	AF416465	AF102780 Banana bu
39	59.6	7.6	1111	14	S56276	AF416465 Banana bu
40	58.6	7.5	1103	6	AR010232	S56276 orf VI (com
41	58.6	7.5	1103	14	AF416469	AR010232 Sequence
42	58.6	7.5	1105	6	AR010234	AF416469 Banana bu
43	58.6	7.5	1105	14	AF416464	AR010234 Sequence
44	58	7.4	1109	6	AR010229	AF416464 Banana bu
45	58	7.4	1109	14	AF416467	AR010229 Sequence
						AF416467 Banana bu

```

stem_loop
40. .70
/notes="stem-loop structure"
103. .975
/notes="ORF 1"
/codon_start=1
/protein_id="AAA42894.1"
/db_xref="GI:323307"
/translation="MGSSIRRCWFTLVNTEEEAANVVRRIESLNLVAIVGDEVAPS
TQQRHLQGFILHKTGRRLQGLKTVLGNDRIHLEPTRGDEQNRDYCSKERVLLHGVP
TPGVKRRFLAQRAFEEDLELDEDPGGRYCVHGAASVETWAAENPFPFVPHNMQ
LEVSAIGEDPADDRITLWICGDDGGKSVAKYLGLPDWFYTCGGTRKDVLYOYLE
DPKRNLIIDVPCRNLEYLNYALLCECVKNRAFSSDIPLSYLGFDHVVHLVFANVLDP
YLKISRDRIKLWNI"
314. .775
/notes="ORF 2"
/codon_start=1
/protein_id="AAA42895.1"
/db_xref="GI:323308"
/translation="MTGFTWSRPVPTNRIETTVNRNGCFSSTESRLVLESKGDHPNP
DLRLNMSAWKQATDEADLYELRWNGQDGLPKIRSHFITIGSLKCLKRESORT
IASSQGYADETETGSPCLPNISDSPTGSTHVEPERTYCTSTRTQNEI"
complement(422. .568)
/notes="ORF 6"
/codon_start=1
/protein_id="AAA42896.1"
/db_xref="GI:323309"
/translation="MEMGTDFORPILSIPPKLRVQRIFGIRLPGGVHQPQQIIVGP
IVAF"
639. .797
/notes="ORF 3"
/codon_start=1
/protein_id="AAA42897.1"
/db_xref="GI:323310"
/translation="MRTRRRRREVRCQISRTQARLVLHMWNQKGRIVPVHRGPKTK
FNPRCTQV"
complement(823. .987)
/notes="ORF 5"
/codon_start=1
/protein_id="AAA42898.1"
/db_xref="GI:323311"
/translation="WTHTLNIPQFYSVPADFQIIRQIDIGKEYMHMVEPKITKGFVFP
RTECPVLNTF"
1098. .1286
/notes="ORF 4"
/codon_start=1
/protein_id="AAA42899.1"
/db_xref="GI:323312"
/translation="MNRVMGGPIKDSIWRTNLLLCLOQTPLSTPIQVSSILEKK
AASLYLPSICFCAIGRLS"
336 a 323 c 300 t
BASE COUNT
ORIGIN

```

	Query Match	88.5%	Score 691	DB 14	Length 1291	
	Best Local Similarity	100.0%	Pred. No. 7.2e-191			
	Matches 691	Conservative 0	Mismatches 0	Indels 0	Gaps 0	
Qy	1	atcggagagccagcgagatcgacaaatcctctggatatcgacagacgagagac	60			
Db	601	ATCGGAGACCCAGCGGACGATCGGACAATCCTCTGGATATCGGACGAGACGAGGAGAC	660			
Qy	61	gggaagtcctggtttggccaaatattctcggactcaagcccgactggtttcacacatgtggt	120			
Db	661	GGGAAGTCCGTGTTGGCAAATATCTCGAGTCAAGCCCGACTGGTCTTACACATGTGGT	720			
Qy	121	ggaaccgaaagagcgattgtaccagtcacatcagagaccacaaacgaaatttaactctc	180			
Db	721	GGAACCGAAGACGATATTGTACCACTACATCGAGGACCCCAACGAAATTTAATCTCTC	780			
Qy	181	gatgtaccagcggtgtaatttgaggtatttaaatattgacctgttagaatgtttaagaac	240			
Db	781	CATGTACCCAGCTGTAATTTAGCTATTTAATTAATATGCCCTGTAGAATGTGTGAAGAAC	840			

```

QY 241 agggcattcagtcgacaaatacgaaccccttagttatcttggttcgacacatgtcat 300
Db 841 AGGGCATTCAGTTCGACAAATACGAACCCCTTAGTTATCTTGGGTTCCGACATGTGCAT 900
QY 301 gtactcgtatttgcgaatgtctcgtcgtgattatttgaataatcagcaggagacagataaaa 360
Db 901 GTACTCGTATTGCGCAATGTCTCGCTGCTGATTTATTTGAAATCAGCAGGAGCAGATAAAA 960
QY 361 ctgtgaatatttaagtattgtgtcatcattacacacatacccccccccacgcgc 420
Db 961 CTGTGGAATATTAAAGTATGTGTCATCTAAATACACCAATACCCGCCGCCACGCGC 1020
QY 421 tatcgtttacattcttatgaataatcctccagcgcgaagcctcgtgaggtactaccggc 480
Db 1021 TATCGTTTACATCTTATCAATATCTGCGCAGGCGGAAGGCTGGAGGTCTACCCGCGC 1080
QY 481 cgaagccggaacaaatgaatcgaatgatttggcgggcccacacataaaagattccattt 540
Db 1081 CGAAGCGCGGAACAAATATGAATCGAGTTATGGCGGGGCCACAAATAAAGATTCCATT 1140
QY 541 ggataagacaaatcgttacttctgttcgtcagtcgacgcaacacactttccacgtcaca 600
Db 1141 CGATAAGAACCAATCTGTTACTTGTCTGCAAGTATGGCGGCGGCAACCACTTCCACGTCAC 1200
QY 601 tccaggtgagtagcttgcggagaaagccgcaagcctctctatcaccgtccatttgtt 660
Db 1201 TCCAGGTGAGTAGCTTGTCTGAGAAAGCCGCAAGCCTCTATCTACCGTCCATTGT 1260
QY 661 ttgcgcgagtcgagcggcgtgagttgatctgg 691
Db 1261 TTTGCGCATCGAGCGGCTGAGTTGATCTGG 1291

RESULT 3
FBL32187 1007 bp DNA circular VRL 16-SBP-2000
LOCUS faba bean necrotic yellows virus C9-Eg gene, isolate Egyptian EVI-93.
ACCESSION AJ132187
VERSION AJ132187.1 GI:4995171
KEYWORDS C9-Eg gene; rep protein.
SOURCE faba bean necrotic yellows virus.
ORGANISM Viruses; ssDNA viruses; Nanovirus.
REFERENCE 1 (bases 1 to 1007)
AUTHORS Katul,L., Timchenko,T., Gronenborn,B. and Vetter,H.J.
TITLE Ten distinct circular ssDNA components, four of which encode putative replication-associated proteins, are associated with the faba bean necrotic yellows virus genome
J. Gen. Virol. 79 (Pt 12), 3101-3109 (1998)
99094637
2 (bases 1 to 1007)
AUTHORS Timchenko,T., de Kouchkovsky,F., Katul,L., David,C., Vetter,H.J. and Gronenborn,B.
TITLE A single rep protein initiates replication of multiple genome components of faba bean necrotic yellows virus, a single-stranded DNA virus of plants
J. Virol. 73 (12), 10173-10182 (1999)
20027244
3 (bases 1 to 1007)
AUTHORS Katul,L.
TITLE Direct Submission
Submitted (09-JAN-1999) Katul L., Inst. f. Pflanzenvirologie, Mikrobiol., u. biol. Sicherheit, Biologische Bundesanstalt f. Land- u. Forstwirtschaft, Messweg 11 -12, 38104 Braunschweig, GERMANY
FEATURES
source Location/Qualifiers
1..1007 /organism="faba bean necrotic yellows virus"
/virus
/isolate="Egyptian EVI-93"
/db_xref="taxon:59817"
/lab_host="Vicia faba"
stem_loop 1..39

```

gene

```

73..1003
/gene="C9-Eg"
CDS
73..918
/gene="C9-Eg"
/function="replication-associated protein"
/notes="32.53 kDa"
/codon_start=1
/evidence=experimental
/product="Rep protein"
/protein_id="CAB44027.1"
/db_xref="GI:4995172"
/db_xref="SPTREMBL:Q9WIK2"
/translation="MSAVNMFVTLNFAPEVPLSFDERVQYAVWQHERVNHDIQGV
QUKKAKMTVNKIIIGGNPHLEKMGSTEEASAYAKBESVRAGWSGELIKGSHK
RKMILIKDPENLEPEPKYRRMAWSAMDESKLAEGGFPYTYLWQETVLGLLE
EPNDRIIIVWYGPNGNEKQSFQKFLGLKDYLYLPGGKTQDMTYLMLKNPKANYVMD
IPRCNSEYLNQFMELIKRNTIFSRYEPVGGIINKIHIVLANVLPDYEKISDRI
KIIYC"
TATA_signal 939..944
/gene="C9-Eg"
polyA_signal 998..1003
/gene="C9-Eg"
BASE COUNT 353 a 161 c 227 g 266 t
ORIGIN

```

Query Match 14.9%; Score 116.4; DB 14; Length 1007;

Best Local Similarity 57.0%; Pred. No. 6.5e-23;

Matches 213; Conservative 0; Mismatches 161; Indels 0; Gaps 0;

```

QY 2 tcggagagcagcggacgacgcacaaatcctctggtatgcggacgagcggagagacg 61
Db 545 TAGAAGAAGAGCCCAATGACCGTATTATTATTGGGTCTACGCCCAATATGTAACGAAG 604
QY 62 ggaagtcctgtttgccaatatctcgactcaagcccgactgttctcacatgtggtg 121
Db 605 GAAATACACAGTTGGTAATCTCTGGGATTAATAAAGATATTACCTTTATTTACCTGGAG 564
QY 122 gaaccagaaagcgtattgttacacgacacacacacacacacacacacacacacac 181
Db 665 GTAAAAACCAAGATATGACATATATGTTAATGAAAAATCAAGGCAATATGTTGATGG 724
QY 182 atgtaccacaggttaatttagagattttaaatatgccctgttagaattgttaagaaca 241
Db 725 ATATCTCTGTTGTAATCTGTAATTTAAATATCAATTTATGGAATTAATTAATAATA 784
QY 242 gggcattcagttcggacaaataacgaaccccttagttatcttgggttcgacacatgtgcatg 301
Db 785 GAACCATATTTAGTTATAAATATGAACCAAGTTGGATGTTATTAATAATAATAACATG 844
QY 302 tactcgtatttcccaatgctcgtcgtgatttataataatcagcaggacagataaaaaac 361
Db 845 TAATTGTATTAGCTAATGATTGCTGATTTATGAAATAAATTAAGTATGAAATAAATA 904
QY 362 tgtggaataatttaa 375
Db 905 TAATTATTGTTAA 918

RESULT 4
LOCUS NYV5966
DEFINITION faba bean necrotic yellows virus C9 gene.
ACCESSION AJ005966
VERSION AJ005966.1 GI:3550532
KEYWORDS C9 gene; component 9; putative; rep protein; replication associated protein.
SOURCE faba bean necrotic yellows virus.
ORGANISM faba bean necrotic yellows virus.
REFERENCE 1 (bases 1 to 1004)
AUTHORS Katul,L.
TITLE Direct Submission

```

JOURNAL	Submitted (06-MAY-1998) Katul L., Institute fuer Biochemie und Pflanzenvirologie, Biologisches Bundesanstalt fuer Land- und Forstwirtschaft, Messeweg 11 - 12, 38104 Braunschweig, GERMANY
REFERENCE	2 (bases 1 to 1004)
TITLE	Katul,L., Timchenko,T., Gronenborn,B. and Vetten,H.J. ren distinct circular ssDNA components, four of which encode putative replication-associated proteins, are associated with the faba bean necrotic yellows virus genome J. Gen. Virol. 79 (Pt 12), 3101-3109 (1998) 99094637
JOURNAL MEDLINE FEATURES	Location/Qualifiers 1..1004 /organism="faba bean necrotic yellows virus" /virion /isolate="SV292-88" /db_xref="taxon:59817" 73..918 /gene="C9" 73..918 /gene="C9" /gene="C9" /function="putative replication associated (rep) protein" /codon_start=1 /product="component 9" /protein_id="CAA06789.1" /db_xref="GI:355033" /db_xref="SPTREMBL:O91252" /translation="MSAVNWVFTLNFAGEVPVLSPDERVOYAVWQHERVNHDIQGV I QLKKKAMNTVKNIGNPPLHKMKSGTSEASAYAQKEESRVAGPWSYGELLKKGSHK RKIMELIKDPENLEEPGKYRRAMAWSAMDSESRKLAEEGFPYAFYSWQETVLGLLEE ENPDRIITWYGPNEGKSQFGFLGKKDLYLPGGKTQDMFTMLMKNPRANVYMD IPNCNELYNOFMELIKNRTIYSYKEPVGCIINNKHIVLVANLPDYEKISQDRI KIIVC"
BASE COUNT	352 a 159 c 226 g 267 t
ORIGIN	
Query Match	14.7%; Score 114.8; DB 14; Length 1004;
Best Local Similarity	56.7%; Pred. No. 1.9e-22;
Matches 212; Conservative	0; Mismatches 162; Indels 0; Gaps 0;
Qy	2 tcggagaccgaggacatcgcaaatcctctggatatgcgacgacgagcggagagacy 61
Db	545 TAGAAGAAGAGCCCAATGACCGCTACTATTATTGGGTGTACGCCCAAAATGGTATGAAG 604
Qy	62 ggaagtccgtgtttgccaaatatctcggaactcaagcccgaactggttctcacatgtggtg 121
Db	605 GRAAATCACAGTTTGSTRAATTCTGGGATTAATAAAAAAGATTACCTTTATTTACCTGGAG 664
Qy	122 gaaccagaagacgctattgtaccagcatcatcgaggagccccaaaataaatttaacctcg 181
Db	665 GTAAAACCCAAGATGACATATATGTTAATGAAAAATCCAAAGGCAGAAATGTTGTGATGG 724
Qy	182 atgtaccagggtgaatttagtagtatttaaatattgcctgttgagaatgtgttaaagaaca 241
Db	725 ATATTCTCGTTGTAATCTGNAATTTAAATTACCAATTTATGGAATTAATTAATAATA 784
Qy	242 gggcatttcagttcggacaaaatacgaaccccttagttatcttgggttcgaccatgtgatcg 301
Db	785 GAACCATATATAGTTATTAATATGAACCAAGTGGATGTATTAAAAATAATAAATAACAAG 844
Qy	302 tactgtatttcgcaatgctcctgcctgattattgaaaaaacagcaggagcagagaataaac 361
Db	845 TAATGTATTAGCTAATGATTTCCTGATTGAAAAAATTAGTACAGGTAGAATTAAAA 904
Qy	362 tgtgggaattattaa 375
Db	905 TAATTTATTGTTAA 918
RESULT 5	
AR063452	
LOCUS	Sequence 4 from patent US 5846705.
DEFINITION	linear DNA 1096 bp PAT 29-SEP-1999

ACCESSION	AR063452	AR063452.1	GI:5992760
VERSION			
KEYWORDS			
SOURCE	Unknown.		
ORGANISM	Unclassified.		
REFERENCE	1 (bases 1 to 1096)		
AUTHORS	Wu,R.-Y., You,L.-R. and Soong,T.-S.		
TITLE	Nucleotide sequence of two circular ssDNA associated with banana bunchy top virus and method for detection of banana bunchy top virus		
JOURNAL	Patent: US 5846705-A 4 08-DEC-1998;		
FEATURES	Location/Qualifiers		
source	1..1096		
BASE COUNT	347 a 231 c 244 g	274 t	
ORIGIN	1..1096		
Query Match	14.6%;	Score 114.2;	DB 6;
Best Local Similarity	57.5%;	Pred. No. 2.9e-22;	
Matches	225;	Conservative 0;	Mismatches 163; Indels 3; Gaps 1;
QY	1	atcgagagccagcgacgacgcacacatcctctgatatcgcgacgacgagagagac	60
Db	551	ATCGAAGGTGTTCTGATGATCGAAGTATCATCTGGGTATACGTCCCAACGGAGCGAA	610
QY	61	ggaaagtcctggtttcccaaatatctcggactcaagccgcactggtctcacacatgtggt	120
Db	611	GGAAAGTCAACCTTCGCAAGATATCTCATTAACCCCGGATGGGATATATCAACGGT	670
QY	121	ggaaccgaaagagcgtattgtaccagtcacatcgagagcccaacgaaatattaatcctc	180
Db	671	GGAAAGCGTCGGATATGATGCACATCATACGATGATCCTGATATCATTCGATTATT	730
QY	181	gatgtaccagcaggtgtaatttagagtatttaaatatccctgttagaagtgttaagaac	240
Db	731	GATATCCCGAAGATCAATCAGATATCTGAATATGGCGTTATAGAACAAATTAAGAT	790
QY	241	agggcattcagttcggacaaatcacgaacc---ccttagttatcttgggttcgacctgtg	297
Db	791	ACAGTTTAAATACAAATACGAACCATGTGTGATTAGAAAAGATGGACAAAATGTC	850
QY	298	catgtactcgtatttgccaatgtcctgcctgattatttgaataatcagcaggagacaata	357
Db	851	CATCTAAATGTTATGGCAAAATGTTGCTGCTGATTATTGTAATAATTCAGAAAGTAGAATA	910
QY	358	aaactgtggaatatttaagtatgtgtcatc	388
Db	911	AAAATAATTAATGTTGAGAAAGGAACATTC	941
RESULT	6		
BYTV2	BYTV2	1096 bp	DNA linear
LOCUS	Banana bunchy top virus (BBTV DNA II) V1, V2, C1 and C2 genes,		
DEFINITION	complete cds's.		
ACCESSION	L32167		
VERSION	L32167.1	GI:520791	
KEYWORDS	stem loop.		
SOURCE	Banana bunchy top virus		
ORGANISM	Banana bunchy top virus		
REFERENCE	1 (bases 1 to 1096)		
AUTHORS	Wu,R.-Y., You,L.-R. and Soong,T.-S.		
TITLE	Nucleotide sequences of two circular single-stranded DNAs associated with banana bunchy top virus		
JOURNAL	Unpublished (1994)		
FEATURES	Location/Qualifiers		
source	1..1096		
	/organism="Banana bunchy top virus"		
	/db_xref="taxon:12585"		
TATA signal	1..7		


```

QY 298 catgtactcgtatttgcgaatgctgcctgctgatttgaataacagcaggacagaata 357
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 856 CATGTAATTCGTAATTCATTCATCCAGCAATTCGTAATAATTCGCCAAGATAGAATA 915

QY 358 aaactgtggaatttaaaagtatgtgcatctaaa 392
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 916 AAGATTATTATTGTTAAATAACACGCTATGACAA 950

RESULT 10
BBU12587 1127 bp DNA circular VRL 01-FEB-1995
LOCUS Banana bunchy top virus DNA IV ORF V1 and ORF C1 genes, complete
DEFINITION
ACCESSION U12587
VERSION U12587.1 GI:642393
KEYWORDS Banana bunchy top virus.
SOURCE Banana bunchy top virus
ORGANISM Viruses; ssDNA viruses; Nanovirus.
REFERENCE 1 (bases 1 to 1127)
AUTHORS Wu, R.-Y. and You, L.-R.
TITLE Nucleotide sequences of DNA III and DNA IV associated with banana bunchy top virus and their relation to other closely related virus DNAs
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1127)
AUTHORS Wu, R.-Y.
TITLE Direct Submission
JOURNAL Submitted (25-JUL-1994) Rey-Yuh Wu, Agricultural Biotechnology Division, Development Center for Biotechnology, 81 Chang Hsing Street, Taipei, 10671, Taiwan, ROC
FEATURES
source
1..1127
/organism="Banana bunchy top virus"
/db_xref="taxon:12585"
/clone="S15 and 4-2"
1..8
15..49
79..933
/note="ORF V1"
/codon_start=1
/protein_id="AAA61877.1"
/db_xref="GI:642394"
/translation="MSSFKWCFTLYNSAAREDFLALLKEELNVAVGVDEVPSSG
QKHLQGLYSLKSIKGLGKKYKSWRAHWRAGSDSDNAKYCSKETLIILELGPASQ
GNSRRKLSEWVSRPERMRLEOPEIYHRTSVAKLKKEEFVHPCLDRPWLQITEA
IDPEPDDRTLFMYGPNNGEKSSTYAKSLMKKDWFTYRGKKENILFVYVDEGSEKHI
VFDPICNQDLYNDVIEALKDRVIESTKYKPIKLVELNNIIVIMANFMPFECKISE
DRIKILYYC"
CDS
complement(497..643)
/note="ORF C1"
/codon_start=1
/protein_id="AAA61878.1"
/db_xref="GI:642395"
/translation="MTSHMLISPHYSDHRRPRRFDHLVPRQLPPSIESAMVYRGKDE
RIPP"
BASE COUNT 353 a 222 c 260 g 292 t
ORIGIN

Query Match 12.6%; Score 98.6; DB 14; Length 1127;
Best Local Similarity 56.4%; Pred. No. 1e-17;
Matches 226; Conservative 0; Mismatches 169; Indels 6; Gaps 2;

QY 1 atcggagagcagcgacatcgacaaatcctctggatctcgacagcagcagcagcagcagc 60
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 559 ATGACGAGGACACAGATGATCGACGACATCTCTGGGTCTATGCTCCGAATGTAATCAG 618

QY 61 ggggaagtcctgtttgcaaatatctcggactcaagccgactggtttctacacatgtggt 120
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 619 GGGAAATCAACATATGCGAAGTCATT---ATGAGGAGGAGCTGGTCTACACGAGAGGT 675

```

```

QY 121 ggaacacaaagagcgctatttgcacagtcacatcgaggaccacaaacgaa---atttaac 177
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 676 GCGAAGAAGAGGAGACATACCTCTCTCTACGTCGACGACGAGGATCTGAGAAGCATATTGTA 735

QY 178 ctcgatgtaccacaggtgaatttagagatttaataattatgccctgttagaattgtttaag 237
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 736 TTGNATATCTCGCTGTAATCAGGATTATTAAATTAATGATGTATAGAGGCATTAAG 795

QY 238 aacaggccattcagttcggacaaatacgaaccccttagttatcttgggttcgaccatgtg 297
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 796 GATAGGGTGATAGAGAGTACATAATAATAAACCTATTAGTTAGTTGAATGAATAATAATA 855

QY 298 catgtactcgtatttgcgaatgctgcctgctgatttgaataacagcaggacagaata 357
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 856 CATGTAATTCGTAATTCATTCATCCAGCAATTCGTAATAATTCGCCAAGATAGAATA 915

QY 358 aaactgtggaatttaaaagtatgtgcatctaaaattacac 398
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 916 AAGATTATTATTGTTAAATAACACGCTATGACAACTGCTAC 956

RESULT 11
AB000922 1000 bp DNA circular VRL 30-OCT-1998
LOCUS Milk vetch dwarf virus genome segment 3 encoding virus
DEFINITION replication-associated protein, complete sequence.
ACCESSION AB000922
VERSION AB000922.1 GI:3808181
KEYWORDS virus replication-associated protein.
SOURCE milk vetch dwarf virus (lab_host:Pisum sativum) DNA.
ORGANISM Viruses; ssDNA viruses; Nanovirus.
REFERENCE 1 (bases 1 to 1000)
AUTHORS Sano, Y.
TITLE Direct Submission
JOURNAL Submitted (08-FEB-1997) Yoshitaka Sano, Kyoto Institute of
Technology, Department of Applied Biology, Matsugasaki, Sakyo-ku,
Kyoto 606-8585, Japan (Tel:075-724-7764, Fax:075-724-7764)
REFERENCE 2 (sites)
AUTHORS Sano, Y., Wada, M., Hashimoto, Y., Matsumoto, T. and Kojima, M.
TITLE Sequences of ten circular ssDNA components associated with the milk
vetch dwarf virus genome
JOURNAL J. Gen. Virol. 79 (Pt 12), 3111-3118 (1998)
MEDLINE 99094638
COMMENT On Oct 29, 1998 this sequence version replaced gi:3798655.
FEATURES
source
1..1000
/organism="milk vetch dwarf virus"
/virion
/db_xref="taxon:67585"
/chromosome="segment 3"
/lab_host="Pisum sativum"
1..29
58..912
/function="helicase"
/note="putative"
/codon_start=1
/product="virus replication-associated protein"
/protein_id="BAA33982.1"
/db_xref="GI:3798656"
/translation="MPTVQSTCWFTLNFKGPILPENVVOYACWOHERVGHDLQ
GFLQAQSLGOMKALIPGAHFEMKRALNSQAKYAMKETRIEGPWEYGVKYYKKG
SHKRIMRFEDDEPMKIEDLSYRLCLSRMTEQSTAEWMDKMPQDVQIQEI
ETDPYRKLIWVYGPKGKGGKSTFYRLSLKGWGLPGPETHMLHTIISGEPKNWV
FDIPRVAEYVNYGVLEQVKNRVNTRYEIVIRDDNHPHVHIVFANCMDFTKISE
DMKIVHC"
TATA_signal 990..996
polyA_signal 333 a 185 c 238 g 244 t
BASE COUNT 333 a 185 c 238 g 244 t
ORIGIN

```


Query Match 11.9%; Score 93; DB 14; Length 1004;
Best Local Similarity 55.3%; Pred. No. 4.5e-16;
Matches 202; Conservative 0; Mismatches 160; Indels 3; Gaps 1;

NRKLLDRYKENPEDELPKARRCRRAKIDKEKFTAEKFEVDEDEBWKILEKEIE
 KIAPRSILWYGPQGGEGTSKAKELITRGKFKDVAISYVEDPTRHVVFD
 IPRMOEYCNYSLEMLKDRILIIISNKYEPIITNCQVNIHVIVMANFLPDVTKISEDRI
 KIIYC"

polyA_signal 348 a 177 c 230 g 252 t
 BASE COUNT
 ORIGIN

Query Match 11.9%; Score 92.8; DB 14; Length 1007;
 Best Local Similarity 54.4%; Pred No. 5, 2e-16;
 Matches 209; Conservative 0; Mismatches 172; Indels 3; Gaps 1;

QY 1 atcgagagcccgagcgacgatcgcaaatcctctcgatatcgcgagcgagcgagagac 60
 DB 565 AITGAAAGATTGCTTCGCCTCGAAGTATCTTTGGGTGTATGGCCCAAGGTGGAGAA 624
 QY 61 gggaaagtcggtgtttgcccataatctcggactcaagcccgactgtttctacacatgtgt 120
 DB 625 GGAAGACGAGTAAAGCAAAAGAACT---ATAACGCGTGGTGGTGGTTTACACGGCGGA 681
 QY 121 ggaaccagaaagagcgtattgaccagtagcagcagcccaaacgaaatttaaccc 180
 DB 682 GGAAGCAAGATGACGCTGGCTTATAGTTATGGAGGACCCCAAGGCATGTGGTATT 741
 QY 181 gatgtaccaggtgttaatttagagtagtttaaatattgcccctgttagaatgttttaagaac 240
 DB 742 GATATACCAAGTGATATGCAAGAAATATTGTAATTACAGCTTAATAGAGATGCTTAAGGAT 801
 QY 241 agggcattcagttcgcaaatcagcaaccccttagttctgttggttcgaccatgtgc 300
 DB 802 AGAATTATTATTAGTAACAATAATATGAGCAATTAACAATTGCCAAGTGTATAATTATCAT 861
 QY 301 gtactcgtatttgcgaatgtccctgcctgattatttgaataatcagcagggagagagataaaa 360
 DB 862 GTAATTGTAATGCTAACTTTTACCTGTATGTAACAAAAATATCTGAAGATAGATAAAAA 921
 QY 361 ctgtggaatatttaagtagtgtgt 384
 DB 922 ATAATTATTGTTGAAATTCGCGT 945

Search completed: June 19, 2002, 15:43:27
 Job time: 13465 sec

QY 12 agcgagcagcacaatcctctcgatatcgagcagagcgagagcgagagtcctgt 71
 DB 551 ACCAGACACAGAACGATAATCTGGTTTACGGATCTGATGGAGGAGAGGTAAATCAAC 610
 QY 72 gttgcaaatatctcgagactcaagcccgactgtttctacacatgtgtggaaccagaaa 131
 DB 611 CTTGCGAAGGAAT---AATTAGTAGTGGTGGTTTATCTCCAGGAGGAAAGCGCA 667
 QY 132 ggaagtattaccagtagatcgagagcccaaacgaaaatttaactctcgatgtaccag 191
 DB 668 GGATATTCTGTATGATGTCTCAAGACCCAGAGAGAAATATCGCTTCGATGTTCCAG 727
 QY 192 gtgtaatttagagtagtttaaatattgcctgttagaatgtgttaagaaacagggcattcag 251
 DB 728 GTGTCTTCGGAGATGATGAATATCAAGCATGGAGATGTTGAAGAACAGAGATTATTCG 787
 QY 252 ttcgacaataacgaaccccttagttatcttgggttcgaccatgtgcattgactcgtatt 311
 DB 788 AAGTACAAATATAGGCCCGTAGATCTTTGCTTAGAAGAAAGCTTCATTAAATTGTT 847
 QY 312 tgcgaatgtcgtcgtatttgaataatcagcagcgagagcagagataaactgtggaatat 371
 DB 848 TGCAAACTGTCACTGACCTGACCCCAAAATAAAGTCAGGACAGAAATTGTAATTCATATG 907
 QY 372 ttaaa 376
 DB 908 TTGAA 912

RESULT 15
 AB000920 1007 bp DNA circular VRL 30-OCT-1998
 LOCUS Milk vetch dwarf virus genome segment 1 encoding viral
 DEFINITION replication-associated protein, complete sequence.
 ACCESSION AB000920
 VERSION AB000920.1 GI:3808179
 KEYWORDS viral replication-associated protein.
 SOURCE milk vetch dwarf virus (lab_host:Pisum sativum) DNA.
 ORGANISM Viruses; ssDNA viruses; Nanovirus.
 REFERENCE 1 (bases 1 to 1007)
 AUTHORS Sano,Y.
 DIRECT SUBMISSION
 TITLE Submitted (08-FEB-1997) Yoshitaka Sano, Kyoto Institute of
 JOURNAL Technology, Department of Applied Biology; Matsugasaki, Sakyo-ku,
 Kyoto 606-8585, Japan (Tel:075-724-7764, Fax:075-724-7764)
 REFERENCE 2 (sites)
 AUTHORS Sano,Y., Wada,M., Hashimoto,Y., Matsumoto,T. and Kojima,M.
 TITLE Sequences of ten circular ssDNA components associated with the milk
 vetch dwarf virus genome
 JOURNAL J. Gen. Virol. 79 (Pt 12), 3111-3118 (1998)
 MEDLINE 99094638
 COMMENT On Oct 29, 1998 this sequence version replaced gi:3798651.
 Sequence updated (30-Sep-1997).
 FEATURES
 source
 1..1007
 /organism="milk vetch dwarf virus"
 /viral
 /db_xref="taxon:67585"
 /chromosome="segment 1"
 /lab_host="Pisum sativum"
 1..33
 51..57
 91..936
 /function="helicase"
 /note="putative"
 /codon_start=1
 /product="viral replication-associated protein"
 /protein_id="BAA33980.1"
 /db_xref="GI:3798652"
 /translation="MPTLQGTWCFTLNFSGDAPSLSPNFVQACWQHERVSHDLQ
 GYIQMKRSTLKMKELLPGAHLEYSKGTPEEASDYAMKEETRVAGPWYTGELLKKG

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2002, 16:50:59 ; Search time 3798.65 Seconds
(without alignments)
352.997 Million cell updates/sec

Title: US-09-462-955B-1_COPY_211_991

Perfect score: 781

Sequence: 1 atcggagacgcagcggacga.....ccccaacctctgtaaacccc 781

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_032802:*

1: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1980.DAT:*

2: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1981.DAT:*

3: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1982.DAT:*

4: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1983.DAT:*

5: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1984.DAT:*

6: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1985.DAT:*

7: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1986.DAT:*

8: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1987.DAT:*

9: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1988.DAT:*

10: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1989.DAT:*

11: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1990.DAT:*

12: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1991.DAT:*

13: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1992.DAT:*

14: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1993.DAT:*

15: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1994.DAT:*

16: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1995.DAT:*

17: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1996.DAT:*

18: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1997.DAT:*

19: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1998.DAT:*

20: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1999.DAT:*

21: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA2000.DAT:*

22: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA2001A.DAT:*

23: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA2001B.DAT:*

24: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	114.2	14.6	1096	AAV71833	BBTV DNA II clone
2	114.2	14.6	1096	AAV38950	Banana bunchy top
3	112.6	14.4	1091	AAV71834	BBTV DNA II clone
4	112.6	14.4	1091	AAV38947	Banana bunchy top
5	112.4	14.4	886	AAV34687	DNA sequence of BB
6	92.6	11.9	1106	AAV71832	BBTV DNA I clone
7	91	11.7	1106	AAV38946	Banana bunchy top
8	89	11.4	1017	AAV13165	SCSV segment 6. S
9	88.4	11.3	1022	AAV13161	SCSV segment 2. S

10	81.2	10.4	593	20	AAV34686	DNA sequence of BB
11	62.8	8.0	1110	19	AAV24089	Banana bunchy top
12	59.6	7.6	982	18	AAV49405	Banana bunchy top
13	59.6	7.6	1110	19	AAV24086	Banana bunchy top
14	59.6	7.6	1111	19	AAV24077	Banana bunchy top
15	59.6	7.6	1111	19	AAV24084	Banana bunchy top
16	59.6	7.6	1111	19	AAV24087	Banana bunchy top
17	58.6	7.5	1103	19	AAV24091	Banana bunchy top
18	58.6	7.5	1105	19	AAV24093	Banana bunchy top
19	58	7.4	1109	19	AAV24088	Banana bunchy top
20	58	7.4	1111	19	AAV24090	Banana bunchy top
21	57	7.3	1104	19	AAV24092	Banana bunchy top
22	56.4	7.2	1111	19	AAV24085	Banana bunchy top
23	42	5.4	300	20	AAV71831	Subgenomic fragmen
24	42	5.4	300	21	AAV38949	Banana bunchy top
25	36.8	4.7	287	20	AAV71830	Subgenomic fragmen
26	36.8	4.7	287	21	AAV38948	Banana bunchy top
27	36.8	4.7	2885	23	AAV70474	DNA encoding novel
28	35	4.5	1935	23	AAV68521	DNA encoding novel
29	35	4.5	1935	23	AAV72032	DNA encoding novel
30	35	4.5	1935	23	AAV73099	DNA encoding novel
31	34.8	4.5	4344	23	AAV73169	DNA encoding novel
32	34.8	4.5	4447	23	AAV72881	DNA encoding novel
33	34.8	4.5	5584	23	AAV69664	DNA encoding novel
34	34.6	4.4	1915	23	AAV81517	DNA encoding novel
35	34.6	4.4	2162	23	AAV76891	DNA encoding novel
36	34.2	4.4	951	23	AAV68084	DNA encoding novel
37	34.2	4.4	2523	23	AAV81683	DNA encoding novel
38	34	4.4	997	23	AAV74263	DNA encoding novel
39	34	4.4	1599	23	AAV73145	DNA encoding novel
40	34	4.4	1599	23	AAV74254	DNA encoding novel
41	34	4.4	2011	23	AAV77251	DNA encoding novel
42	34	4.4	3489	23	AAV68093	DNA encoding novel
43	34	4.4	4331	23	AAV72876	DNA encoding novel
c 44	33.8	4.3	1182	23	AAV81606	DNA encoding novel
c 45	33.6	4.3	1479	23	AAV92524	DNA encoding novel

ALIGNMENTS

RESULT 1

AAV71833	10-FEB-1999 (first entry)
ID	AAV71833 standard; DNA; 1096 BP.
XX	AC
XX	AAV71833;
XX	AC
XX	10-FEB-1999 (first entry)
DT	BBTV DNA II clone (2-17) nucleotide sequence.
XX	BBTV DNA II clone (2-17) nucleotide sequence.
DE	Banana bunchy top virus; BBTV; PCR technique; plant; tissue; banana;
XX	Musa; phloem damage; aphid; nucleic acid detection; circular; cyclic; ss.
OS	Banana bunchy top virus.
XX	XX
FH	Key
FT	TATA_signal
FT	Location/Qualifiers
FT	1..7
FT	/*tag= a
FT	stem_loop
FT	8..38
FT	/*tag= b
FT	CDS
FT	70..928
FT	/*tag= c
FT	/*product= "ORF-V2 product"
FT	/transl_except= (pos:215..217, aa:Gly)
FT	533..538
FT	/*tag= d
FT	polyA_signal
FT	799..804
FT	/*tag= e
FT	polyA_signal
FT	907..912
FT	/*tag= f
FT	polyA_signal
FT	1030..1035
FT	/*tag= ge

[illegible]

KEY	RESULT	6
AAV71832	AAV71832 standard; DNA; 1106 BP.	
AAV71832;		
10-FEB-1999	(first entry)	
BBTV DNA I clone (7-4-2)	nucleotide sequence.	
Banana bunchy top virus; BBTV; PCR technique; plant; tissue; banana; Musa; phloem damage; aphid; nucleic acid detection; circular; cyclic; ss.		
Banana bunchy top virus.		
Key	Location/Qualifiers	
TATA_signal	1..8	
stem_loop	/*tag= a	
	9..38	
	/*tag= b	
CDS	62..922	
	/*tag= c	
	/product= "ORF-V2 product"	
	/transl_except= (pos:335..337, aa:Gly)	
	/transl_except= (pos:518..520, aa:Ser)	
polyA_signal	380..385	
	/*tag= d	
polyA_signal	901..906	
	/*tag= e	
US5846705-A.		
08-DEC-1998.		
06-APR-1995;	95US-0418071.	
06-APR-1995;	95US-0418071.	
(BIOT-)	DEV CENT BIOTECHNOLOGY.	
Soong T, Wu R, You L;		
WPI; 1999-059037/05.		
P-PSDB; AAW87459.		
Nucleic acids having banana bunchy top virus component sequences - used to design primers for use in polymerase chain reaction detection of the virus		
Claim 1; Fig 11A-B; 27pp; English.		
This represents the nucleotide sequence of a banana bunchy top virus (BBTV) DNA I (clone 7-4-2) circular single stranded (css) DNA. The invention provides nucleic acid sequences associated with BBTV that can be used in a PCR technique for detecting BBTV. The nucleic acid		

[illegible]

Nucleic acids having banana bunchy top virus component sequences -
used to design primers for use in polymerase chain reaction
detection of the virus

Claim 1; Fig 11A-B; 27pp; English.

This represents the nucleotide sequence of a banana bunchy top virus
(BBTV) DNA I (clone 7-4-2) circular single stranded (css) DNA. The
invention provides nucleic acid sequences associated with BBTV that can
be used in a PCR technique for detecting BBTV. The nucleic acid

[illegible]

Db 790 aggtattattaaattcatggtttattagagggaatttaagaatlggaataattcaaaagcgagg 849
Qy 260 aatacgaaccccttagttatcttgggttcgaccatgtgcatgtactcggtatttgccaatg 319
Db 850 aatacgaacccgttttgaagatag--tgaataatgtcgaagtcattgtgaatggctaact 906
Qy 320 tcctgcctgattatttgaataatcagcaggggacagagaataaaactg 363
Db 907 tccttcggaagggaaggaatctttctgaagatcgaataaagt 950

Search completed: June 19, 2002, 16:51:06
Job time: 17524 sec

QY 241 agggcattcagtcggaacatacaaccccttagttatttctggttcgacattgcat 300
Db 841 agggcattcagtcggaacatacaaccccttagttatttctggttcgacattgcat 900
QY 301 gctactcgtatttgcaatgctcgtcgtgatttattgaaatacagcagggacagataaaa 360
Db 901 gctactcgtatttgcaatgctcgtcgtgatttattgaaatacagcagggacagataaaa 960
QY 361 ctgtggaatatttaagtagtctgcatctaatattacacataccgcccgcacgcgc 420
Db 961 ctgtggaatatttaagtagtctgcatctaatattacacataccgcccgcacgcgc 1020
QY 421 tatcgtttacattcttaataatctgcccagcgcgaagcgtgaggtgctacccgcgc 480
Db 1021 tatcgtttacattcttaataatctgcccagcgcgaagcgtgaggtgctacccgcgc 1080
QY 481 cgaagccgggaacataatgaatcagttatggcgggcccaataaaagattccattt 540
Db 1081 cgaagccgggaacataatgaatcagttatggcgggcccaataaaagattccattt 1140
QY 541 ggataagacgaatctgttacttctgctgagtcgacgcaacacatttccacgtcaccaa 600
Db 1141 ggataagacgaatctgttacttctgctgagtcgacgcaacacatttccacgtcaccaa 1200
QY 601 tccagtgtagtctgctgaggaagaaagcgcgaagcctctatctacccatttgg 660
Db 1201 tccagtgtagtctgctgaggaagaaagcgcgaagcctctatctacccatttgg 1260
QY 661 ttgcgcgacgcgagcgcgtgagttgattgg 691
Db 1261 ttgcgcgacgcgagcgcgtgagttgattgg 1291

RESULT 2

US-08-418-071-4
; Sequence 4, Application US/08418071
; Patent No. 5846705
; GENERAL INFORMATION:
; APPLICANT: Wu, Rey-Yuh
; APPLICANT: You, Li-Ru
; APPLICANT: Soong, Tai-Seng
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCE OF TWO CIRCULAR ssDNA
; TITLE OF INVENTION: ASSOCIATED WITH BANANA BUNCHY TOP VIRUS AND METHOD FOR DETECTI
; TITLE OF INVENTION: BANANA BUNCHY TOP VIRUS
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Neave
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 5.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/418,071
; FILING DATE: 06-APR-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley Jr., James F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: DCB-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 596-9000
; TELEFAX: (212) 596-9090
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1096 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single

; TOPOLOGY: circular
; MOLECULE TYPE: genomic DNA
; DESCRIPTION: /desc "BBTV DNA II (clone 2-17)"
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Banana Bunchy Top Virus (BBTV)
; US-08-418-071-4

Query Match 14.6%; Score 114.2; DB 2; Length 1096;
Best Local Similarity 57.5%; Pred. No. 6.6e-30; Indels 3; Gaps 1;
Matches 225; Conservative 0; Mismatches 163;
QY 1 atcggagagcagcggcagcgcacacatctctggtatctggtatcgagcagcagcagcagc 60
Db 551 ATCGAAGGTGTTCTGATGATCGAAGTATCATCTGGGTATACGTCCTCCACGAGCGAA 610
QY 61 gggagtcgctgtttggccaaatctctggactcaagcccgactgggtctacacatgtgtt 120
Db 611 GGAAGTCAACCTTCGCAAGATATCTATCATTAACCCGGATGGGATATATCAACGGT 670
QY 121 ggaacgaagaagcagctattgtaccagtcacatcgagcaccacaaagaaatttaactc 180
Db 671 GGAAGACGTCGGGATATGTCACATCATTAACGATGATCTGATATCATTAATGATT 730
QY 181 gatgtaccagctgttaatttagagatttaaatatccctgttagaattgttaagaac 240
Db 731 GATATCCCGCAAGTCATCTCAGATTATCTGAATTTATGGCGTTATAGAACAAATTAAGA 790
QY 241 agggcattcagtcggaacatacaaccccttagttatttctggttcgacattgcat 297
Db 791 AGAGTTTAAATAAATACAAAATACGAACCATGTGTGATTAGAAAAGATGGACAAATGTC 850
QY 298 catgtactcgtatttgccaatgctcctgctgatttattgaaatacagcagggacagata 357
Db 851 CATGTAATTGTTATGGCAAAATGCTTGCTGATATTGTTAAATTTTCAGAAGATAGAATA 910
QY 358 aaactgtggaatatttaagtagtctgcatc 388
Db 911 AAAATAATTAATTGTTGAGAAAGGAAACTTC 941

RESULT 3

US-08-418-071-5
; Sequence 5, Application US/08418071
; Patent No. 5846705
; GENERAL INFORMATION:
; APPLICANT: Wu, Rey-Yuh
; APPLICANT: You, Li-Ru
; APPLICANT: Soong, Tai-Seng
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCE OF TWO CIRCULAR ssDNA
; TITLE OF INVENTION: ASSOCIATED WITH BANANA BUNCHY TOP VIRUS
; TITLE OF INVENTION: BANANA BUNCHY TOP VIRUS
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Neave
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 5.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/418,071
; FILING DATE: 06-APR-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley Jr., James F.

DETE

[illegible]

Query Match	11.3%;	Score 88.4;	DB 4;	Length 1022;
Best Local Similarity	54.3%;	Pred. No. 6.8e-21;		
Matches 201;	Conservative	0;	Mismatches 166;	Indels 3; Gaps 1;
QY				
DB				
7	gagccagcgacgacgacacaaatcctctggatgatacgagcagagacgagagacgggaag	66		
559	GCGGAACGATGATCGGAGTATCATCTGGGTCTATGGTTTCAGACGGAGGACGAAGAAG	618		
67	tccgtgtttccaaaatctcggactcaagcccgactggtttctacacatggtgtgaaacc	126		
QY				
619	ACGAGCTTCGCGAAGGAATT---AATCAGTAGTGATGCTTTTATACAGCCGAGGGAAG	675		
DB				
127	agaaggaagctattgtaccagtagacatcgaggaccccaaaacgaatttaactcgcagtga	186		
QY				
676	ACCAGGAGCGTATTATATATGATGTCTCAAGACCCAGAGAGGAATAATCGGTTGTATGTT	735		
DB				
187	cccaggtgtaatttagagtattaaattatgccctgttagaattgtttaagacacgggca	246		
QY				
736	CCCAGGTGTTCTCGGAGATGAATATCAGGCGATGGAGATGTTGAAGAACAGAGTT	795		
DB				
247	ttcagttcgcgacaaatcgaacccttagttatctctgggttcgaccatbtgcatctc	306		
QY				
796	TTTGCAGTACAAATATATAGCCCTGTAGACTCTTTGTATTAGGAAGTAGTTCATTTAATT	855		
DB				
307	gtatttgcgaatgtcctgcctgattatttgaataatcagcaggggacagaataaaactgtgg	366		
QY				


```

RESULT 10
US-08-202-186-9
; Sequence 9, Application US/08202186
; Patent No. 5756708
;
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequencing
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:

```

```

RESULT 11
US-08-202-186-12
; Sequence 12, Application US/08202186
; Patent No. 5756708
;
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner

```


Thu Jun 20 06:56:35 2002

ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W.
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/202,186
FILING DATE: 24-FEB-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: JEFFERY, Donald D.
REGISTRATION NUMBER: 19,980
REFERENCE/DOCKET NUMBER: 71611/102 FIKE
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202 672 5300
TELEFAX: 202 672 5399
TELEX: 904136

INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 1103 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: circular

US-08-202-186-16

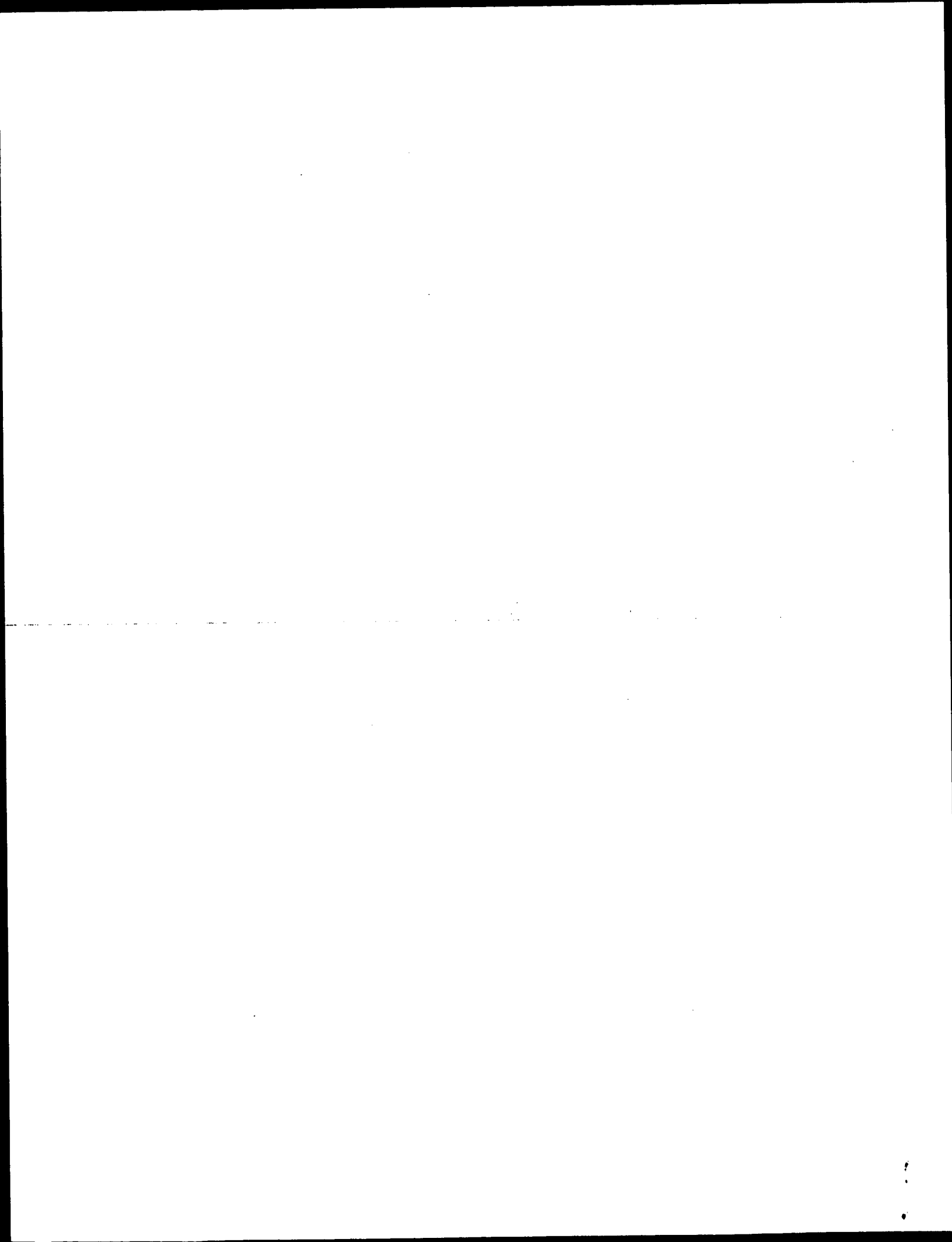
Query Match	7.5%;	Score 58.6;	DB 1;	Length 1103;
Best Local Similarity	52.5%;	Pred. No. 2e-10;		
Matches 180;	Conservative 0;	Mismatches 154;	Indels 9;	Gaps 2;
Qy	20	atcgcaaatcctctggatctcgacagacgagagacgagacggaatcgctgtttgccca	79	
Db	616	ATCGAGGAAATATTGGGTCTACGCCCAATAATGAGCGAAGNAGACAACTTTGCCAA	675	
Qy	80	aatatctcggaactcaagccgactgtttctcacatgtgttgaaacagaaagacgtat	139	
Db	676	ACATTTAATGAGACTAAGATGCGTTTATTCGCCAGGAGGAAATCATTTGGATATAT	735	
Qy	140	tgtaccagtcatactgagagaccacaacaaatttaactctcgtatgtaccagctgtaatt	199	
Db	736	GTAGATTGTATAATTATGAGGATATA-----GTTATATTGTATTTCCAGATGCAAG	789	
Qy	200	tagagtatttaaatatgcctgttagaatgtgttaagaacaggccattcagttcggaca	259	
Db	790	AGGAATATTTAAACTATGCTTATTAGAAGAAATTTAAATGGAATTTATTCAAAGCGGA	849	
Qy	260	aatacgaaccccttagttatcttgggttcgaccatgctatgtactcgtattgccaatg	319	
Db	850	AATATGAACCCGTTTGAAATG--TAGAATATGTGGAAGTCATTGTAATGCTAACT	906	
Qy	320	tcctgcctgattattgaaatacagcaggagacagataaaaact	362	
Db	907	TCCPTCCGAAGGAAGAACTCTTTCTCGAAGATCGAATAAAGCT	949	

RESULT 14
US-08-202-186-18
; Sequence 18, Application US/08202186
; Patent No. 5756708
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
;

STREET: 3000 K Street, N.W.
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/202,186
FILING DATE: 24-FEB-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: JEFFERY, Donald D.
REGISTRATION NUMBER: 19,980
REFERENCE/DOCKET NUMBER: 71611/102 FIXE
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202 672 5300
TELEFAX: 202 672 5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 1105 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: circular
J5-08-202-186-18

Query Match	7.5%; Score 58.6; DB 1; Length 1105;
Best Local Similarity	52.5%; Pred. No. 2e-10;
Matches	180; Conservative 0; Mismatches 154; Indels 9; Gaps
Qy 20	atcgacaaatcctctggatatcggaacagacgaggaagcgcggtgtttgcc 79
Db 618	ACCGAAGGATAAATTGGGTCTATGGCCCCAATGGAGGAGAGGAACAACTATGCA 677
Qy 80	aatatctcgactcaagccgactggtctcacatgtggtggaaccagaagcgctat 139
Db 678	AATATTTAATGNAGACGAAGATGCTTTATTGCCAGGAGGAAAAATCATTTGGATATAT 737
Qy 140	tgtaccagtcatactcgaggaccocaaacgaaatttaatcctcgatgataccagggtgaatt 199
Db 738	GFAGATTGTTATAATTATAGGAAATA-----GTTATATTTTGATATCCCGATCGAAAG 791
Qy 200	tagagtatttaaatatgcctctgttagaatgtgttaagaacagggcattcagttcggaca 259
Db 792	AGGAATATTTAAACATATGTTTATTAGAGAAATTTAAGAAATCGAATTATTCAAAGCGGA 851
Qy 260	aatacgaaccctctagttatcttgggttcgaccatgtgcattcgtactcgtatttggcaatg 319
Db 852	AATATGAACCCGTTTTGAAAATTG---TAGAATATGTGGAAGTCATTGTAAATGGCTAACT 908
Qy 320	tctgcgctgattatttgaaaatcagcggggacagaataaaact 362
Db 908	TCCCTCCGAGCAGCAACATCTTTTCAGAAAGATCGAAGCTAAAGCT 951

RESULT 15
US-08-202-186-13
; Sequence 13, Application US/08202186
; Patent No. 5756708
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2002, 14:04:12 ; Search time 7489.97 Seconds
(without alignments)
1407.365 Million cell updates/sec

Title: US-09-462-955B-1_COPY_211_991

Perfect score: 781

Sequence: 1 atcgagagccgagcga.....ccccaaacctgtgtaacccc 781

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
C 1	41	5.2	745	9	AV908194 AV908194
C 2	40	5.1	512	12	AQ121628 HS_3089_A
C 3	38	4.9	2275	9	AF034173 AF034173
4	37.8	4.8	706	9	AL669587 AL669587
5	36.8	4.7	498	12	AQ615572 HS_5144_B
6	36.8	4.7	518	9	AW818171 CM1-ST027
7	36.4	4.7	787	12	AL098749 Drosophil
C 8	36	4.6	507	12	A2444169 Drosophil
C 9	36	4.6	929	12	BH137680 ENPFC13TR
C 10	35.6	4.6	318	12	AQ089903 HS_3001_A
11	35.6	4.6	1101	12	CNS016H0
12	35.4	4.5	460	9	A1942816 fc67b05.x
13	35.4	4.5	488	12	AL085215 Arabidops
C 14	35.2	4.5	345	12	A2401897
15	35.2	4.5	848	12	AQ746046 HS_2277_A
16	35	4.5	524	12	BH175883 004_P_17-
17	35	4.5	524	12	AL612823 T3 end of

18	34.6	4.4	487	12	AQ418525
19	34.6	4.4	555	12	AQ417598
20	34.6	4.4	665	9	BB633121
C 21	34.6	4.4	689	12	AG167308
C 22	34.4	4.4	594	12	AZ397924
C 23	34.4	4.4	898	9	AL667528
24	34.4	4.4	905	10	B1957755
C 25	34	4.4	442	12	AQ228962
C 26	34	4.4	443	12	AQ010299
C 27	34	4.4	1101	12	CNS00DBS
C 28	33.8	4.3	410	12	AQ080176
C 29	33.8	4.3	604	12	AG019820
C 30	33.8	4.3	673	12	AG167164
C 31	33.8	4.3	749	12	AZ331069
C 32	33.8	4.3	772	12	AQ489882
C 33	33.8	4.3	865	12	BH135588
C 34	33.8	4.3	873	12	CNS012PA
35	33.8	4.3	894	12	CNS032CO
C 36	33.8	4.3	1174	10	BE779656
C 37	33.6	4.3	326	12	AQ008775
C 38	33.6	4.3	403	10	BG006561
C 39	33.6	4.3	418	12	AQ632834
40	33.6	4.3	434	12	AQ728714
C 41	33.6	4.3	605	12	AQ481591
C 42	33.6	4.3	658	12	AQ077737
C 43	33.6	4.3	974	10	BG787937
44	33.4	4.3	283	9	AI065300
C 45	33.4	4.3	348	12	AQ348508

ALIGNMENTS

RESULT AV908194/c
LOCUS AV908194 745 bp mRNA linear EST 09-NOV-2001
DEFINITION AV908194 Nori Satoh unpublished cDNA library, young adult Ciona intestinalis cDNA clone rciad75n04 3', mRNA sequence.
ACCESSION AV908194
VERSION AV908194.1 GI:16897292
KEYWORDS EST.
SOURCE Ciona intestinalis.
ORGANISM Ciona intestinalis.
REFERENCE 1 (bases 1 to 745)
AUTHORS Satoh,N., Satou,Y., Kohara,Y. and Shin-I,T.
TITLE Expressed genes in Ciona intestinalis
JOURNAL Unpublished (2000)
COMMENT Contact: Nori Satoh
Department of Zoology
Kyoto University
Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
Tel: 81-75-753-4081
Fax: 81-75-705-1113
Email: satoh@ascidian.zool.kyoto-u.ac.jp.

FEATURES

source

Location/Qualifiers
1..745
/organism="Ciona intestinalis"
/db_xref="taxon:7719"
/clone="rciad75n04"
/clone_lib="Nori Satoh unpublished cDNA library, young adult"
/tissue_type="whole animal"
/dev_stage="young adult"
239 a 121 c 139 g 244 t 2 others

BASE COUNT 239 a 121 c 139 g 244 t 2 others
ORIGIN

Query Match 5.2%; Score 41; DB 9; Length 745;
Best Local Similarity 47.5%; Pred. No. 0.12;
Matches 122; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

```

Qy 103 tgggtctacacatgtgtggaaccagagaaagtgattgtaccaggtacatcaggagccca 162
Db 644 TGGTTCACAAATCCACAGCCAGATTAAAGCTCAAGTAATCAAGTCCCACAGGATCCA 585

Qy 163 aaacgaaatttaactctcgatgtaccaggtgttaatttagagattttaaattatgcctg 222
Db 584 AAGTGTGTTCTCATCAAAATATAACTACAAAGTGCATAAATAACTGTACAATGTTGCTGCATTT 525

Qy 223 ttagaatgtgtaagacacagggaattcagttcggacaaataacgaacccttaattatctt 282
Db 524 ATAGTTCCTAAATAGGACTAAGGCATGATGATCTATTATATATATCCACACACTGCAATGACTT 465

Qy 283 gggttcgacctgtgcattactgtatttgccaaatgtcctgcctgattatttgaaatc 342
Db 464 GTGCTGTACAGTTTCCATACCTTGCCATTGGCAGTTGAGTGTAGTAATTAATCATATTTAACA 405

Qy 343 agcaggggacagataaa 359
Db 404 AACTTTTAGCAAAAAA 388

RESULT 2
A121628/c
LOCUS
DEFINITION
  HS_3089_Al_B03_MF CIT Approved Human Genomic Sperm Library D Homo
  sapiens genomic clone Plate=3089 Col=5 Row=C, DNA sequence.
ACCESSION
  A121628
VERSION
  A121628.1 GI:3498794
KEYWORDS
  GSS.
SOURCE
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  1 (bases 1 to 512)
  Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
  Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
  Hood,L.
  Sequence-tagged connectors: A sequence approach to mapping and
  scanning the human genome
  Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
JOURNAL
  99380589
MEDLINE
  Contact: Mahairas GG, Wallace JC, Hood L
  High Throughput Sequencing Center
  University of Washington
  401 Queen Anne Avenue North, Seattle, WA 98109, USA
  Tel: (206) 616-3618
  Fax: (206) 616-3887
  Email: jwallace@u.washington.edu
  Sequence Tagged Connector
  Plate: 3089 row: C column: 5
  Class: BAC ends
  High quality sequence stop: 512.
FEATURES
  source
    Location/Qualifiers
      1..512
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="Plate=3089 Col=5 Row=C"
        /clone_lib="CIT Approved Human Genomic Sperm Library D"
        /sex="male"
        /note="organ: sperm; Vector: pBeloBAC11; BAC Clones In
          E-Coli DH10B"
BASE COUNT 174 a 85 c 115 g 136 t 2 others
ORIGIN

Query Match 5.1%; Score 40; DB 12; Length 512;
Best Local Similarity 59.8%; Pred. No. 0.21;
Matches 67; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

Qy 278 atcttggttcgaccatgtgcattactgtatttgccaatgtcctgctgattattga 337
Db 179 ATCCTTATTGCCCCATCTGCTTTACTCCACTTTACCAATGCTCTTTTAAAGTATCCCA 120

```

```

Qy 338 aaatcagcaggagacagaataaaactgtggaatattttaaagtattgtgtcatct 389
Db 119 AACTGAACATAGATATGATCAGACTCAGAGAGATTAAATGAAGACACATTT 68

RESULT 3
AF034173
LOCUS
DEFINITION
  AF034173 Human mRNA (Tripodis and Ragoussis) Homo sapiens cDNA
  clone ntcon2 contig, mRNA sequence.
ACCESSION
  AF034173
VERSION
  AF034173.1 GI:2707735
KEYWORDS
  EST.
SOURCE
  human.
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  1 (bases 1 to 2275)
  Tripodis,N. and Ragoussis,J.
  Generation of a transcription map in the region immediately
  centromeric to human MHC across the 6p21.2-6p21.3 chromosomal
  boundary
  Unpublished (1997)
JOURNAL
  Contact: Tripodis, Nikos
  Division of Medical and Molecular Genetics
  Guys Hospital
  7th floor, Guy's Tower, London SE1 9RT, UK
  Email: nikos@nki.ni.
FEATURES
  Location/Qualifiers
    1..2275
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /map="6p21.3"
      /clone="ntcon2 contig"
      /clone_lib="Human mRNA (Tripodis and Ragoussis)"
BASE COUNT 438 a 619 c 470 g 599 t 149 others
ORIGIN

Query Match 4.9%; Score 38; DB 9; Length 2275;
Best Local Similarity 16.7%; Pred. No. 1.9;
Matches 32; Conservative 85; Mismatches 75; Indels 0; Gaps 0;

Qy 352 agataaaactgtggaattatttaaagtattgtcatctaaataaccacacacccgccg 411
Db 1469 AAAAAAAMWRYKRWKRKRRTGTGMYKRYRAMMAMCAMACWYWKMRGKKC 1528

Qy 412 ccacgcgtctgtttacattctatgaatctcctgccagcgcgaagcctggggagtg 471
Db 1529 WKYKRYKRYTSTYKSWRWYWTYYTYWCCTSMKSASCAMRWGMYGSRSSRSYW 1588

Qy 472 ctaccggccgaagccgggaacaataatgaatcgagttatggcgccgcccacataaaag 531
Db 1589 GYWGMSGCGYGMTKRYRYSYTGTGWTWTTTWWYKMSMTWTTTWTWTRTKTWWWWW 1648

Qy 532 attccatttga 543
Db 1649 WTCWTMRKRGA 1660

RESULT 4
AL669587
LOCUS
DEFINITION
  AL669587 directional larval cDNA library Ciona intestinalis cDNA
  clone 050ZA01 5', mRNA sequence.
ACCESSION
  AL669587
VERSION
  AL669587.1 GI:18142845
KEYWORDS
  EST.
SOURCE
  Ciona intestinalis.
  Ciona intestinalis
  Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
  Phlebobranchia; Clonidae; Ciona.
REFERENCE
  1 (bases 1 to 706)

```


QY	260	aatacgaaccccttagtattcggttcgaccatgcgatcatctactcgtatttgccaatg.	319
Dd	726	dggcgcagacakykadtgatATGRKARSGAGGKKTGRTGTTRTARTATWGGDRKGAW	785
QY	320	tectgcgtgattttttaaaatcacgccaggacagaataaaaactgtggaaatttaaatga	379
Dd	786	TTCCTKTTRKANCKTDRAKTGARTGAGGWBGARAVTGWGWTATWDABSTIDWRGAG	845
QY	380	tgtgtcaatcaattaa 395 : : :::: :	
Dd	846	NRTAATTAATTWTVA 861 : : :::: :	
RESULT	12		
A1942816			
LOCUS	fcb7b05_xl Zebrafish WashU MPIMG EST Danio rerio cDNA clone	mRNA linear	EST 07-JUN-2001
DEFINITION	IMAGE:3726417 3', mRNA sequence.		
ACCESSION	A1942816		
VERSION	A1942816.1 GI:5707472		
SOURCE	zebrafish.		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio. 1 (bases 1 to 460)		
REFERENCE	Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy, .S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood, .K., Steptoe,M., Theising,B., Allen.M., Bowers,X., Person.B., Swaller,T., Gibbons M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston.R., and Wilson,K.		
TITLE	WashU Zebrafish EST Project 1998		
JOURNAL	Unpublished (1998)		
COMMENT	Contact: Stephen L. Johnson Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810 Email: zbrafish@watson.wustl.edu cDNA Library Preparation: Matthew Clark. cDNA Library Arrayed by: Matthew Clark. DNA Sequencing by: Washington University Genome Sequencing Center Clone Distribution: Genome Systems, St. Louis, Missouri (web address: www.genomesystems.com) (email contact: info@genomesystems.com) and Research Genetics, Huntsville, Alabama (web address: www.resgen.com) (email contact: info@resgen.com) and Ressourcenzentrum Primardatenbank, Berlin, Germany (web address: www.rzpd.de) Seq primer: T7 ET from Amersharn High quality sequence stop: 443. Location/Qualifiers 1..460 /organism="Danio rerio" /db_xref=taxon:"7955" /cclone_lib="Zebrafish WashU MPIMG EST" /sext="mixed" /tissue.type="26 somite embryos, adult livers, shield stage embryos" /lab_host="Xtl-blue VRF"		
FEATURES	source		
	/note=Vector: pSPORNI; Site_1: NotI; Site_2: Salt; Ist strand cDNA was primed with a Not I.; oligo(drf)y15 primer [5'pgACTAGTTCTAGATCGCAGGCCGCCTTTTTTTTTTTT3']; double-stranded cDNA was ligated to Sal I adaptors (BRL), digested with Not I and cloned into the Not I and Sal I sites of the pSPORN vector (BRL). Library was constructed by Matthew Clark (Lehrach lab: ICRRF, London and Max Planck Institut fuer Molekulare Genetik, Berlin). cDNAs for EST analysis were selected following oligonucleotide hybridization fingerprinting of arrayed clones from zebrafish late somitogenesis (26 ss), adult liver or embryonic shield stage (5.6 h) libraries.fingerprint		

[illegible]

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2002, 15:43:27 ; Search time 5904.86 Seconds
(without alignments)
2066.124 Million cell updates/sec

Title: US-09-462-955B-1_COPY_409_991

Perfect score: 583

Sequence: 1 ttagagtatttaattatgc.....ccccaaacctgtgtaacccc 583

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vl.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vi.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	Match	Length	ID	Description
No.	Score	Match	Length	ID	Description

1	493	84.6	1291	6	AR172770	AR172770 Sequence
2	493	84.6	1291	14	CFDCG	M29963 Coconut fol
3	68.4	11.7	1007	14	FBE132187	AJ132187 faba bean
4	65.2	11.2	1004	14	NYV5966	AJ005966 faba bean
5	59.8	10.3	1111	14	BBU12586	U12586 Banana bunc
6	59.4	10.2	1127	14	BBU12587	U12587 Banana bunc
7	58	9.9	1106	6	AR063451	AR063451 Sequence
8	58	9.9	1106	14	BYTV1	L32166 Banana bunc
9	57.8	9.9	1095	14	U02312	U02312 Banana bunc
10	57.4	9.8	1109	14	AF216221	AF216221 Banana bu
11	52.2	9.0	1091	6	AR063453	AR063453 Sequence
12	52.2	9.0	1096	6	AR063452	AR063452 Sequence
13	52.2	9.0	1096	14	BYTV2	L32167 Banana bunc
14	52	8.9	1095	14	AF216222	AF216222 Banana bu
15	49.6	8.5	1007	14	AB000920	AB000920 Milk vetc
16	46	7.9	1000	14	AB000922	AB000922 Milk vetc
17	45.6	7.8	1017	6	ARI45398	ARI45398 Sequence
18	45.6	7.8	1017	14	SCU16735	UI6735 Subterranea
19	45	7.7	1015	14	FBE132185	AJ132185 faba bean
20	41.8	7.2	1343	8	ACY14168	Y14168 Ageratum co
21	41.4	7.1	1367	14	AYE238493	AJ238493 Nanovirus
22	41	7.0	1014	14	NYV5964	AJ005964 faba bean
23	40.2	6.9	151886	9	AC022335	AC022335 Homo sapi
24	40	6.9	26445	9	AC084298	AC084298 Homo sapi
25	40	6.9	125020	9	AF429315	AF429315 Homo sapi
26	39.8	6.8	439	8	AF158545	AF158545 Pinus tae
27	38.6	6.6	1022	14	AB009047	AB009047 Milk vetc
28	38.6	6.6	118595	9	AC002379	AC002379 Human HAC
29	37.4	6.4	1004	14	NYV5968	AJ005968 faba bean
30	37	6.4	16473	5	AP002927	AP002927 Polymixia
31	37	6.3	210115	9	AL442127	AL442127 Human DNA
32	36.8	6.3	36652	9	AC005176	AC005176 Homo sapi
33	36.8	6.3	55801	2	AC087476	AC087476 Homo sapi
34	36.8	6.3	102456	9	AP000083	AP000083 Homo sapi
35	36.8	6.3	140770	2	AF252832	AF252832 Homo sapi
36	36.8	6.3	152812	2	AC090536	AC090536 Homo sapi
37	36.8	6.3	153684	2	AC013645	AC013645 Homo sapi
38	36.8	6.3	160556	2	AC013645	AC013645 Homo sapi
39	36.8	6.3	170414	2	AC026019	AC026019 Homo sapi
40	36.8	6.3	172820	2	AC092899	AC092899 Homo sapi
41	36.8	6.3	176730	9	AC055839	AC055839 Homo sapi
42	36.8	6.3	182529	2	AC064832	AC064832 Homo sapi
43	36.8	6.3	187927	9	AC046142	AC046142 Homo sapi
44	36.6	6.3	136232	2	AC024006	AC024006 Homo sapi
45	36.6	6.3	142555	9	AL139214	AL139214 Human DNA

ALIGNMENTS

RESULT	1	AR172770	1291 bp	DNA	linear	PAT 17-DEC-2001
LOCUS	AR172770	Sequence 1 from patent US 6303345.				
DEFINITION	AR172770	Sequence 1 from patent US 6303345.				
ACCESSION	AR172770	Sequence 1 from patent US 6303345.				
VERSION	AR172770.1	GI:17912261				
KEYWORDS	Unknown.					
SOURCE	Unknown.					
ORGANISM	Unknown.					
REFERENCE	1 (bases 1 to 1291)					
AUTHORS	Rohde,W., Becker,D., Randles,J.W., Hehn,A. and Salamini,F.					
TITLE	Use of a virus DNA as promoter					
JOURNAL	Patent: US 6303345-A 1 16-OCT-2001;					
FEATURES	Location/Qualifiers					
source	1..1291					
BASE COUNT	336 a	323 c	332 g	300 t		
ORIGIN						

Query Match 84.6%; Score 493; DB 6; Length 1291;
Best Local Similarity 100.0%; Pred. No. 1.6e-134;

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source

1. (bases 1 to 1095)
Yeh,H.H., Su,H.J. and Chao,Y.
Genome characterization and identification of viral-associated
dsDNA component of banana bunchy top virus
Virology 198, 645-652 (1994)
94120739
2. (bases 1 to 1095)
Chao,Y.
Direct Submission
Submitted (04-OCT-1993) Yu-Chan Chao, Institute of Molecular
Biology, 120, Sec. 2., Yen-joe-yuan Rd., Nan-Kang, Taipei, Taiwan
115, Republic of China
Location/Qualifiers
1. .1095

BASE	COUNT	ORIGIN
337	a	217 c 253 g 288 t
		QTLEAIDEEDPDRSLIIVVGPYGNEGK"
		/translation="MIARSPDRKKEIQPEIFHRYOSVKNLKKFEEFVHFCLODRPQWI
		/db_xref="GI:475435"
		/protein_id="AA17784.1"
		/codon_start=1
		/note="orf2"
		419 634
		"
		TGQKHLQGLYSLLKMKPRIEEVLPGSLGDCQNRREEFVLFQRPNPRIIVSCC
		/translation="MSPSLKWCFTINSSAAERENFLSLKEEDVHYAVVGDEVAPA
		/db_xref="GI:475434"
		/protein_id="AA17783.1"
		/codon_start=1
		/note="orf1"
		82 . .390
		/notes="component 2 (BETV-C2)"
		/clone="pETd18"
		/db_xref="taxon:12585"
		/specific_host="Musa acuminata Colla"
		/organism="Banana bunchy top virus"

Query Match	9.9%	Score	57.8	DB	14	Length	1095;
Best Local Similarity	55.8%	Pred. No.	5e-06;				
Matches	110;	Conservative	0;	Mismatches	87;	Indels	0; Gaps
Qy	4	gagtatataattgcccgttagaattgtttaagacacgggcatttcagttcgacaaa	63				
Db	767	GATTATTAAATTATGATCTAATAGAGCCATTAAAGGATAGGGCTTATAGAGAGTACTAAA	826				
Qy	64	tacgaaccccttagttatttgggttcgaccatgcatgtactctatttggcaatgtc	123				
Db	827	TACAAACCCTAATAAGATAGTTGAATTAGGTAAAAATACATTAATCTCATGCGCAATTTC	886				
Qy	124	ctgccgtattatttgaataatcagcgggacagataaaactctggaataatttaaagtatg	183				
Db	887	ATGCCGTACTCTGTAAATCTCCGAAGATCGATAAATAATCATTTATTGCTGAATAACA	946				
Qy	184	tgtcatctaaattacac	200				
Db	947	CGTATGACAATCGTAC	963				

RESULT	10
AF216221	
LOCUS	Banana bunchy top virus satellite SI replication initiation protein
DEFINITION	AF216221 1109 bp DNA linear VRL 08-MAR-2001 Banana bunchy top virus satellite SI replication initiation protein (ORF V1) gene, complete cds.
ACCESSION	AF216221
VERSION	AF216221.1 GI:12004325
KEYWORDS	
SOURCE	Banana bunchy top virus.
ORGANISM	Banana bunchy top virus
REFERENCE	Viruses; ssDNA viruses; Nanovirus. 1 (bases 1 to 1109) Horser,C., Harding,R. and Dale,J. Banana bunchy top nanovirus DNA-1 encodes the 'master' replication
AUTHORS	
TITLE	

JOURNAL	initiation protein
MEDLINE	J. Gen. Virol. 82 (Pt 2), 459-464 (2001)
PUBMED	21102983
REFERENCE	11161286
AUTHORS	2 (bases 1 to 1109)
TITLE	Horser,C.L., Karan,M., Harding,R.M. and Dale,J.L. Additional rep-encoding DNAs associated with banana bunchy top virus
JOURNAL	Arch. Virol. 146 (1), 71-86 (2001)
REFERENCE	3 (bases 1 to 1109)
AUTHORS	Horser,C.L., Karan,M., Harding,R.M. and Dale,J.L.
TITLE	Direct Submission
JOURNAL	Submitted (14-DEC-1999) School of Life Sciences, Queensland University of Technology, GPO Box 2434, Brisbane, QLD 4001, Australia
FEATURES	Location/Qualifiers
source	1..1109 /organism="Banana bunchy top virus" /isolate="Taiwan" /db_xref="taxon:12585" /note="satellite S1" 3..33
stem loop	

[illegible]

	Query Match	9.8%	Score 57.4;	DB 14;	Length 1109;
	Best Local Similarity	54.5%	Pred. No. 6.5e-06;		
	Matches 115;	Conservative	0;	Mismatches 96;	Indels 0;
QY	4	gagttattaaattatgccccctgttagaattgtttaagaacacagggcattcagttcgacaaa	63		
Db	744	GATTATTTAAATTATGATGTTATAGAGCATTAAAGGATAGGGTGATAGACAGTACTATAA	803		
QY	64	tacgaaccccttagttcttgggttcgacattgtcattctcgttatttgcgaatgtc	123		
Db	804	TATAAACCTATTAAGTTAGTTGAATTGATTAAATATACATGTTAATGTTCATGGCTAAATTC	863		
QY	124	ctgcctgattatttggaaaaatcagcaggacagataaaaactgtggaattatttaagtatg	183		
Db	864	ATGCCAGAAATTCGTAAAACTCCGGAAGATAGATAAAGATTATTATTGTGTAAAAAGGA	923		
QY	184	tgtcatctaaattacaccaataccgcgcgcg	214		
Db	924	AATTATATATTATGCACTATATGACAAATTCGTACGC	954		

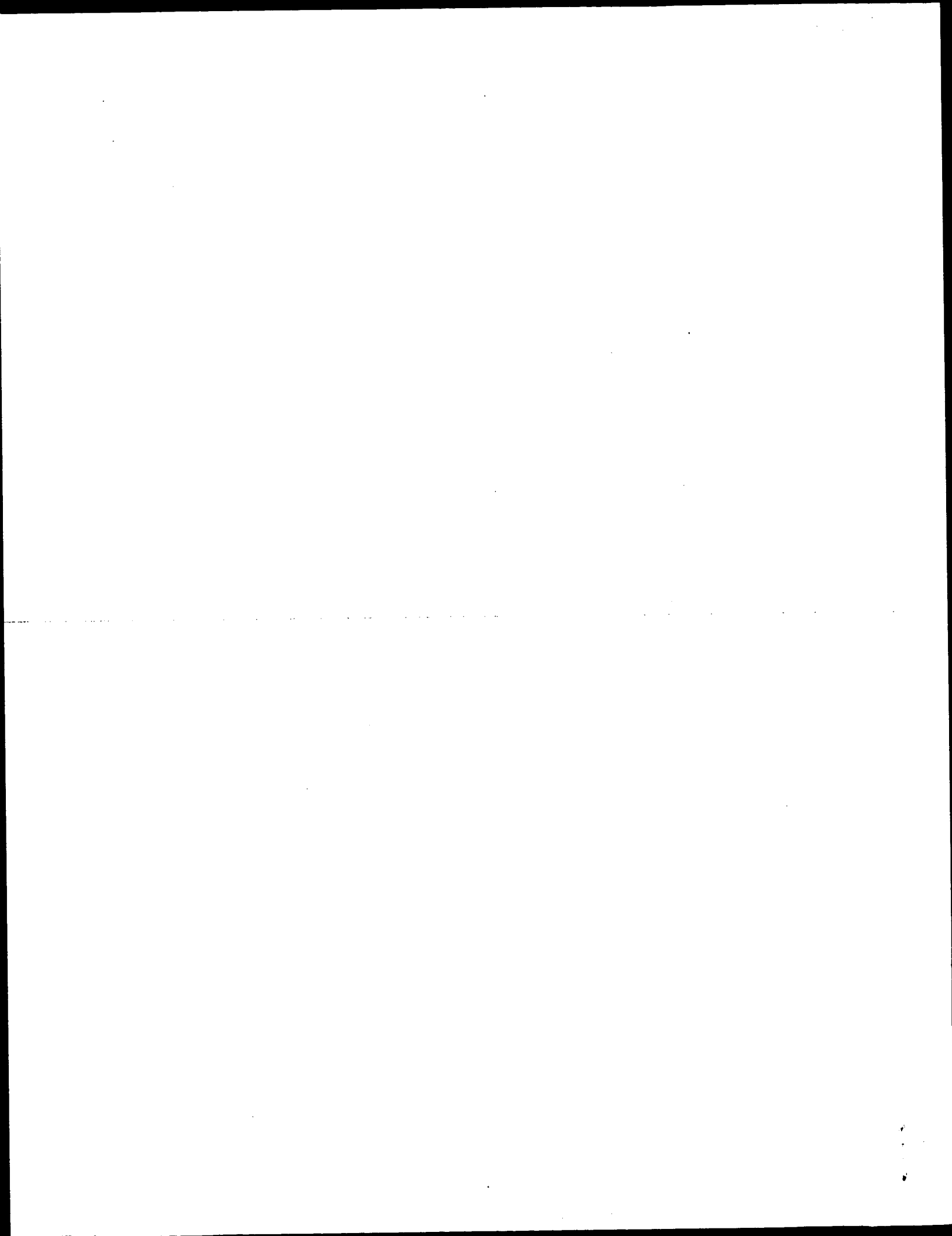
[illegible]

Db 762 AGAATATTGTAATTACAGCTTAATAGAGATCGCTTAAGGATAGAATTATTATTAGTAAACA 821

Db 762 AGAATATTGTAATTACAGCTTAATAGAGATCGCTTAAGGATAGAATTATTATTAGTAAACA 821

QY	63	atcgaacccctctagttatcttgggttcgacacatgtgcacatgactcgtatttgcgaatgt	122
Db	822	ATATGAGCGCAATTACAAATGCCAAGTCTATAATTTCATGTAATGTGAATGCTAACTT	881
QY	123	cctgcctgattattgaaatcagcaggcagacgaataaaactgtggaattttaagaat	182
Db	802	TTTACCTGATGTACACAAAATATCTGAAGATAGAAATAAAATAATTTATTGTTGAAATTC	941
QY	183	gtgt	186
Db	942	CGGT	945

Search completed: June 19, 2002, 15:43:34
Job time: 13472 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2002, 16:51:06 ; Search time 3798.65 Seconds
(without alignments)
263.504 Million cell updates/sec

Title: US-09-462-955b-1_COPY_409_991
Perfect score: 583
Sequence: 1 tttagtatttaattatgc.....cccccaacctgtgtaacccc 583

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1980.DAT:*
- 2: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1981.DAT:*
- 3: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1982.DAT:*
- 4: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1983.DAT:*
- 5: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1984.DAT:*
- 6: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1985.DAT:*
- 7: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1986.DAT:*
- 8: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1987.DAT:*
- 9: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1988.DAT:*
- 10: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1989.DAT:*
- 11: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1990.DAT:*
- 12: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT:*
- 13: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1992.DAT:*
- 14: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1993.DAT:*
- 15: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1994.DAT:*
- 16: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1995.DAT:*
- 17: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT:*
- 18: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT:*
- 19: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT:*
- 20: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT:*
- 21: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT:*
- 22: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:*
- 23: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT:*
- 24: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	58	9.9	1106	20	AAV71832
2	58	9.9	1106	21	AAV38946
3	57.4	9.8	593	20	AAA34686
4	52.2	9.0	1091	20	AAV71834
5	52.2	9.0	1091	21	AAA38947
6	52.2	9.0	1096	20	AAV71833
7	52.2	9.0	1096	21	AAA38950
8	52	8.9	886	20	AAA34687
9	45.6	7.8	1017	17	AAT13165

10	36.8	6.3	2885	23	AA570474	DNA encoding novel
11	35	6.0	1935	23	AA568521	DNA encoding novel
12	35	6.0	1935	23	AA572032	DNA encoding novel
13	35	6.0	1935	23	AA573099	DNA encoding novel
14	34.6	5.9	1103	19	AAV24091	Banana bunchy top
15	34.6	5.9	1105	19	AAV24093	Banana bunchy top
16	34.6	5.9	1915	23	AA581517	DNA encoding novel
17	34.6	5.9	2162	23	AA576891	DNA encoding novel
18	34.2	5.9	951	23	AA568084	DNA encoding novel
19	34.2	5.9	2523	23	AA581683	DNA encoding novel
20	34	5.8	997	23	AA574263	DNA encoding novel
21	34	5.8	1110	19	AAV24089	Banana bunchy top
22	34	5.8	1111	19	AAV24087	Banana bunchy top
23	34	5.8	1599	23	AA573145	DNA encoding novel
24	34	5.8	1599	23	AA574254	DNA encoding novel
25	34	5.8	2011	23	AA577251	DNA encoding novel
26	34	5.8	3489	23	AA568093	DNA encoding novel
27	34	5.8	4331	23	AA572876	DNA encoding novel
28	33.8	5.8	1182	23	AA581606	DNA encoding novel
29	33.6	5.8	1479	23	AA592524	DNA encoding novel
30	33.6	5.8	1638	23	AA568724	DNA encoding novel
31	33.6	5.8	2610	23	AA583845	DNA encoding novel
32	33.6	5.8	42488	22	AAK66772	Human immune/haema
33	33.2	5.7	1011	23	AA574571	DNA encoding novel
34	33	5.7	1104	19	AAV24092	Banana bunchy top
35	33	5.7	9345	21	AAZ93966	Mouse uromodulin p
36	32.6	5.6	2313	23	AA574278	DNA encoding novel
37	32.6	5.6	2313	23	AA579107	DNA encoding novel
38	32.6	5.6	2313	23	AA592935	DNA encoding novel
39	32.6	5.6	2414	23	AA584293	DNA encoding novel
40	32.6	5.6	3368	23	AA568231	DNA encoding novel
41	32.4	5.6	982	18	AAT49405	Banana bunchy top
42	32.4	5.6	1109	19	AAV24088	Banana bunchy top
43	32.4	5.6	1111	19	AAV24077	Banana bunchy top
44	32.4	5.6	1111	19	AAV24084	Banana bunchy top
45	32.4	5.6	2207	23	AA571746	DNA encoding novel

ALIGNMENTS

RESULT 1
AAV71832
ID AAV71832 standard; DNA; 1106 BP.

XX AC AAV71832;

XX DT 10-FEB-1999 (first entry)

XX DE BBTV DNA I clone (7-4-2) nucleotide sequence.

XX DE Banana bunchy top virus; BBTV; PCR technique; plant; tissue; banana;
KW Musa; phloem damage; aphid; nucleic acid detection; circular; cyclic; ss.

XX OS Banana bunchy top virus.

XX FH Key Location/Qualifiers

FT TATA_signal 1..8

FT stem_loop /*tag= a

FT CDS /*tag= b

FT /*tag= c

FT /product= "ORF-V2 product"

FT /transl_except= (pos:335..337, aa:Gly)

FT /transl_except= (pos:518..520, aa:Ser)

FT polyA_signal 380..385

FT /*tag= d

FT polyA_signal 901..906

FT /*tag= e

XX US5846705-A.

XX

CC to detect BBTV infection.. The PCR technique is used for detecting BBTV
CC in plant tissues (preferably banana, especially Musa species). The
CC virus, one of the most important banana species viruses, causes phloem
CC damage and is transmitted by aphids. PCR detection gives accurate,
CC reliable and specific determination of absence or presence of the
CC virus.
XX
SQ Sequence 1091 BP; 343 A; 224 C; 248 G; 276 T; 0 other:

Query Match 9.0%; Score 52.2; DB 20; Length 1091;
Best Local Similarity 58.0%; Pred. No. 2.5e-07;
Matches 112; Conservative 0; Mismatches 78; Indels 3.

QY	1	ttagagttattaaattgcctgttgaattgtttaaacacaggagcattcaagtctcgac	60
Db	744	tcagattctcgaattatggcgttatagaacaattlaagaatagagttttaataataca	80
QY	61	aaatacgaacc---ccttagttatcttgggtctgacacatgtgcattactcgtattgccc	117
Db	804	aaatacgaaccatgtgattagaaaagatggacaataatgcattgttatggca	86
QY	118	aatgtcctgcctgattattgaaatacagcaggacagataaaaaactgtggaattttaa	177
Db	864	aatgtgttgctgattattgaaaatttcagaagatagataaaaaataattcaattgtga	923
QY	178	agtatgtgtcatc	190
Db	924	gaagggaaacttc	936

RESULT 5
AAA38947
ID AAA38947 standard; DNA: 1091 BP.

XX

DT 23-AUG-2000 (first entry)

Banana bunchy top virus related nucleotide sequence #2.

KW Banana bunchy top virus; BBTV; detection; ds.

TW360710-A.

11-JUN-1999.

30-JUN-1994; 94TW-0106105.

30-JUN-1994; 94TW-0106105.

PA (BIOT-) DEV CENT BIOTECHNOLOGY.

Wu R, You L, Song T:

WPI; 2000-316145/27.

Two circular single-stranded DNAs associated with banana bunchy top virus and detection of the virus -

Claim 2; Page 1; 7pp; Chinese.

The present invention describes two circular single-stranded DNAs associated with banana bunchy top virus (BBTV) and the encoded proteins. Also described is a method of detecting the BBTV virus using the polymerase chain reaction (PCR). The present sequence represents a BBTV related nucleotide sequence from the present invention.

Q Sequence 1091 BP; 345 A; 224 C; 247 G; 275 T; 0 other;

Query Match 9.0%; Score 52.2; DB 21; Length 1091;
 Best Local Similarity 58.0%; Pred. No. 2.5e-07;
 Matches 112; Conservative 0; Mismatches 78; Indels 3; Gaps 1;

QY 1 tttagagtatttaattgcctgttagaattgtttaagaacagggcattcagttcggac 60
 Db 744 tcagattatcgtgaattatggcgttatagaacaaattaaagaatttttaataataaca 803

QY 61 aaatacgaacc---ccttagttatcttgggttcgacatgtgcattcgtctattgcc 117
 Db 804 aaatacgaaccatgtgtgattagaagaatggacaaatgtccatgttaattgtatggca 863

QY 118 aatgtcctgcctgattattgaaataacagcagggacagataaaactgtggaattttaa 177
 Db 864 aatgtgttcctgattattgaaatttcagaagatagaataaaataataattgttga 923

QY 178 agtatgtgtcatc 190
 Db 924 gaaaggaaacttc 936

RESULT 6
 AAV71833
 ID AAV71833 standard; DNA; 1096 BP.
 AC AAV71833;
 XX
 XX
 XX 10-FEB-1999 (first entry)
 DE BBTV DNA II clone (2-17) nucleotide sequence.
 XX
 XX Banana bunchy top virus; BBTV; PCR technique; plant; tissue; banana;
 KW Musa; phloem damage; aphid; nucleic acid detection; circular; cyclic; ss.
 XX
 XX Banana bunchy top virus.
 XX
 XX Key Location/Qualifiers
 FT TATA_signal 1..7
 FT stem_loop 8..38 /*tag= a
 FT CDS 70..928 /*tag= b
 FT /*tag= c
 FT /product= "ORF-V2 product"
 FT /transl_except= (pos:215..217, aa:Gly)
 FT polyA_signal 533..538 /*tag= d
 FT polyA_signal 799..804 /*tag= e
 FT polyA_signal 907..912 /*tag= f
 FT polyA_signal 1030..1035 /*tag= ge
 FT
 FT US5846705-A.
 PN
 PN 08-DEC-1998.
 PD
 PD 06-APR-1995; 95US-0418071.
 PF
 PF 06-APR-1995; 95US-0418071.
 PR
 PR (BIOT-) DEV CENT BIOTECHNOLOGY.
 XX
 XX Soong T, Wu R, You L;
 PI
 PI WPI; 1999-059037/05.
 DR
 DR P-PSDB; AAW87460.
 XX
 XX Nucleic acids having banana bunchy top virus component sequences -
 PT used to design primers for use in polymerase chain reaction
 PT detection of the virus
 XX

PS Claim 2; Columns 17-18; 27pp; English.
 XX
 CC This represents the nucleotide sequence of a banana bunchy top virus
 CC (BBTV) DNA II (clone 2-17) circular single stranded (css) DNA. The
 CC invention provides nucleic acid sequences associated with BBTV that can
 CC be used in a PCR technique for detecting BBTV. The nucleic acid
 CC sequences (AAV71830 to AAV71833) are used as the basis for the
 CC construction of PCR primers, to detect BBTV infection. The PCR technique
 CC is used for detecting BBTV in plant tissues (preferably banana,
 CC especially Musa species). The virus, one of the most important banana
 CC species viruses, causes phloem damage and is transmitted by aphids. PCR
 CC detection gives accurate, reliable and specific determination of absence
 CC or presence of the virus.
 XX
 SQ Sequence 1096 BP; 347 A; 231 C; 244 G; 274 T; 0 other;

Query Match 9.0%; Score 52.2; DB 20; Length 1096;
 Best Local Similarity 58.0%; Pred. No. 2.5e-07;
 Matches 112; Conservative 0; Mismatches 78; Indels 3; Gaps 1;

QY 1 tttagagtatttaattgcctgttagaattgtttaagaacagggcattcagttcggac 60
 Db 749 tcagattatcgtgaattatggcgttatagaacaaattaaagaatttttaataataaca 808

QY 61 aaatacgaacc---ccttagttatcttgggttcgacatgtgcattcgtctattgcc 117
 Db 809 aaatacgaaccatgtgtgattagaagaatggacaaatgtccatgttaattgtatggca 868

QY 118 aatgtcctgcctgattattgaaataacagcagggacagataaaactgtggaattttaa 177
 Db 869 aatgtgttcctgattattgaaatttcagaagatagaataaaataataattgttga 928

QY 178 agtatgtgtcatc 190
 Db 929 gaaaggaaacttc 941

RESULT 7
 AAA38950
 ID AAA38950 standard; DNA; 1096 BP.
 XX
 XX AAA38950;
 XX
 XX 23-AUG-2000 (first entry)
 DT
 DE Banana bunchy top virus related nucleotide sequence #5.
 DE Banana bunchy top virus; BBTV; detection; ds.
 XX
 XX Banana bunchy top virus.
 XX
 XX TW360710-A.
 PN
 PN 11-JUN-1999.
 PD
 PD 30-JUN-1994; 94TW-0106105.
 PF
 PF 30-JUN-1994; 94TW-0106105.
 PR
 PR (BIOT-) DEV CENT BIOTECHNOLOGY.
 XX
 XX Wu R, You L, Song T;
 PI
 PI WPI; 2000-316145/27.
 DR
 DR Two circular single-stranded DNAs associated with banana bunchy top
 PT virus and detection of the virus -
 PT
 XX Claim 2; Page 1; 7pp; Chinese.
 PS
 XX The present invention describes two circular single-stranded DNAs
 CC associated with banana bunchy top virus (BBTV) and the encoded proteins.
 CC

Query Match 7.8%; Score 45.6; DB 17; Length 1017;
 Best Local Similarity 56.4%; Pred. No. 4.6e-05;
 Matches 106; Conservative 0; Mismatches 79; Indels 3; Gaps 1;

QY 1 tttagatatttaattgacctgttagaattgtttaagaacagggcattcagttcgac 60
 Db 726 tcagagtagtgatgattggttaagaacaggttaagaattagggttaaggtaatact 785

QY 61 aaatacgaacc---ccttagttatcttggttcgacatgtgcattgtactctgtattgcc 117
 Db 786 agtatgagccatgtgtaattgcggatgataatactcctgttcattgattgttgca 845

QY 118 aatgctcctgctgatttattgaaatacagcagggagacagataaaactgtggaattattaa 177
 Db 846 aatgactcctcagatttgggaaataagtagaagataaataaatttaattcgtgttga 905

QY 178 agtatgtg 185
 Db 906 aaactctg 913

RESULT 10
 AAS70474
 ID AAS70474 standard; cDNA; 2885 BP.
 XX
 AC AAS70474;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE DNA encoding novel human diagnostic protein #6278.
 XX
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 KW
 XX Homo sapiens.
 OS
 XX WO200175067-A2.
 PN
 XX 11-OCT-2001.
 PD
 XX 30-MAR-2001; 2001WO-US08631.
 XX
 XX 31-MAR-2000; 2000US-0540217.
 PR
 XX 23-AUG-2000; 2000US-0649167.
 PR
 XX (HYSE-) HYSEQ INC.
 PA
 XX Drmanac RT, Liu C, Tang YT;
 PI
 XX WPI: 2001-639362/73.
 DR
 XX P-PSDB; ABG06287.
 XX
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity
 XX
 XX Claim 1; SEQ ID No 6278; 103pp; English.
 PS
 XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC inaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 2885 BP; 1144 A; 623 C; 512 G; 606 T; 0 other;

Query Match 6.3%; Score 36.8; DB 23; Length 2885;
 Best Local Similarity 56.7%; Pred. No. 0.092;
 Matches 68; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 34 gtttaagaacagggcattcagttcgacacaaatacgaaccccttagttattcgttgcgac 93
 Db 1333 gttctggcagggaatcaggaagacaaataaagggtatttaagttaggaaaagag 1392

QY 94 catgtgcattgtactgtatttgcacatgtcctgcctgatttggaaatcagcagggac 153
 Db 1393 gaagtcaattgtcctgttgcagatgacatgattgtttatttagaaatccccatcgac 1452

RESULT 11
 AAS68521
 ID AAS68521 standard; cDNA; 1935 BP.
 XX
 AC AAS68521;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE DNA encoding novel human diagnostic protein #4325.
 XX
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 KW
 XX Homo sapiens.
 OS
 XX WO200175067-A2.
 PN
 XX 11-OCT-2001.
 PD
 XX 30-MAR-2001; 2001WO-US08631.
 XX
 XX 31-MAR-2000; 2000US-0540217.
 PR
 XX 23-AUG-2000; 2000US-0649167.
 PR
 XX (HYSE-) HYSEQ INC.
 PA
 XX Drmanac RT, Liu C, Tang YT;
 PI
 XX WPI: 2001-639362/73.
 DR
 XX P-PSDB; ABG04334.
 XX
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity
 XX
 XX Claim 1; SEQ ID No 4325; 103pp; English.
 PS
 XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC inaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.

CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 1935 BP; 802 A; 414 C; 325 G; 394 T; 0 other;

Query Match 6.0%; Score 35; DB 23; Length 1935;
 Best Local Similarity 56.5%; Pred. No. 0.31;
 Matches 65; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
 QY 34 gtttaagaacaggcatttcagttcgacaaatacgaaccccttagttatcttgggttcgac 93
 ||| | ||||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 1411 gttctggccaggcgaatcgaaggcgagagaagaataaagggtattcattaggaagaag 1470
 QY 94 catgtgcattgtactcgtatttgcgaatgtcctgcctgattatttgaataatcagca 148
 ||| | | | ||||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 1471 gaagtcaaatgtccctgtttgcagatgacatgactgtattatttgaataacccca 1525

RESULT 12
 AAS72032
 ID AAS72032 standard; cDNA; 1935 BP.
 XX
 AC AAS72032;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE DNA encoding novel human diagnostic protein #7836.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX Homo sapiens.
 XX WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI; 2001-639362/73.
 DR P-PSDB; ABG07845.

XX New isolated polynucleotide and encoded polypeptides, useful in
 XX diagnostics, forensics, gene mapping, identification of mutations
 XX responsible for genetic disorders or other traits and to assess
 XX biodiversity -
 XX
 PS Claim 1; SEQ ID No 7836; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques

CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 1935 BP; 802 A; 414 C; 325 G; 394 T; 0 other;

Query Match 6.0%; Score 35; DB 23; Length 1935;
 Best Local Similarity 56.5%; Pred. No. 0.31;
 Matches 65; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
 QY 34 gtttaagaacaggcatttcagttcgacaaatacgaaccccttagttatcttgggttcgac 93
 ||| | ||||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 1411 gttctggccaggcgaatcgaaggcgagagaagaataaagggtattcattaggaagaag 1470
 QY 94 catgtgcattgtactcgtatttgcgaatgtcctgcctgattatttgaataatcagca 148
 ||| | | | ||||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 1471 gaagtcaaatgtccctgtttgcagatgacatgactgtattatttgaataacccca 1525

RESULT 13
 AAS73099/c
 ID AAS73099 standard; cDNA; 1935 BP.
 XX
 AC AAS73099;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE DNA encoding novel human diagnostic protein #8903.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX Homo sapiens.
 XX WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI; 2001-639362/73.
 DR P-PSDB; ABG08912.

XX New isolated polynucleotide and encoded polypeptides, useful in
 XX diagnostics, forensics, gene mapping, identification of mutations
 XX responsible for genetic disorders or other traits and to assess
 XX biodiversity -
 XX
 PS Claim 1; SEQ ID No 8903; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome

CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 1935 BP; 394 A; 325 C; 414 G; 802 T; 0 other;

Query Match 6.0%; Score 35; DB 23; Length 1935;
 Best Local Similarity 56.5%; Pred. No. 0.31;
 Matches 65; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
 QY 34 gtttaagaacaggcattcagttcgacaaatacgaaccccttagttatcttggttcgac 93
 DB 525 GTTCTGCCAGGCAATCATCGGGGAGAAAGAAATAAAGGGTATTCATTAGGAAAGAG 466
 QY 94 catgtcgatctactgtatttgcgaatgctcctgcctgatttgaataatcagca 148
 DB 465 GAAGTCAATTGTCCTGTTGCAGATGACATGATATTTAGAAAACCCCA 411

RESULT 14
 AAV24091
 ID AAV24091 standard; DNA; 1103 BP.
 XX
 AC AAV24091;
 XX
 DT 11-AUG-1998 (first entry)
 DE Banana bunchy top virus component 1 DNA sequence.
 DE BBTV; probe; diagnostic primer; component 1; ss.
 KW Banana bunchy top virus.
 XX OS US5756708-A.
 XX PN 26-MAY-1998.
 XX PF 24-FEB-1994; 94US-0202186.
 XX PR 24-FEB-1994; 94US-0202186.
 XX PA (UYQU-) UNIV QUEENSLAND TECHNOLOGY.
 XX PI Burns TM, Dale JL, Harding RM, Karan M;
 XX WPI; 1998-321636/28.
 XX Isolated Banana bunchy top virus DNA - useful as diagnostic probes
 XX and primers for producing virus-resistant plants
 XX Disclosure; Column 25-28; 59pp; English.
 XX This sequence represents a DNA fragment from the component 1 of
 CC the Banana bunchy top virus (BBTV). The DNA molecules can be used as
 CC diagnostic probes or primers or can be inserted into plants or other
 CC organisms e.g. to produce virus-resistant plants or to act as promoters,
 CC enhancers or termination signals.

XX
 SQ Sequence 1103 BP; 355 A; 192 C; 277 G; 279 T; 0 other;
 Query Match 5.9%; Score 34.6; DB 19; Length 1103;
 Best Local Similarity 55.3%; Pred. No. 0.32;
 Matches 89; Conservative 0; Mismatches 69; Indels 3; Gaps 1;
 QY 4 gagtatttaattatgcctgttagaattgtttaagaacaggcattcagttcgacaaa 63
 DB 792 gaattatttaactatgcttattagaagaatttaaaatgaattattcaaacgaggaaa 851
 QY 64 taagaaccccttagttatcttgggttgcaccatgctgattctgatttgcgaatgct 123
 DB 852 tatgaacccgtttgaaaattg---tagaattgtgaagtcattgttaattgctaaattc 908
 QY 124 ctgctgattattgaaaatcagcaggagacagaataaaact 164
 DB 909 ctccgaaggaaggaattctttctgaagatcgataaagct 949

RESULT 15
 AAV24093
 ID AAV24093 standard; DNA; 1105 BP.
 XX
 AC AAV24093;
 XX
 DT 11-AUG-1998 (first entry)
 DE Banana bunchy top virus component 1 DNA sequence.
 DE BBTV; probe; diagnostic primer; component 1; ss.
 KW Banana bunchy top virus.
 XX OS US5756708-A.
 XX PN 26-MAY-1998.
 XX PF 24-FEB-1994; 94US-0202186.
 XX PR 24-FEB-1994; 94US-0202186.
 XX PA (UYQU-) UNIV QUEENSLAND TECHNOLOGY.
 XX PI Burns TM, Dale JL, Harding RM, Karan M;
 XX WPI; 1998-321636/28.
 XX Isolated Banana bunchy top virus DNA - useful as diagnostic probes
 XX and primers for producing virus-resistant plants
 XX Disclosure; Column 29-30; 59pp; English.
 XX This sequence represents a DNA fragment from the component 1 of
 CC the Banana bunchy top virus (BBTV). The DNA molecules can be used as
 CC diagnostic probes or primers or can be inserted into plants or other
 CC organisms e.g. to produce virus-resistant plants or to act as promoters,
 CC enhancers or termination signals.
 XX
 SQ Sequence 1105 BP; 359 A; 197 C; 279 G; 270 T; 0 other;

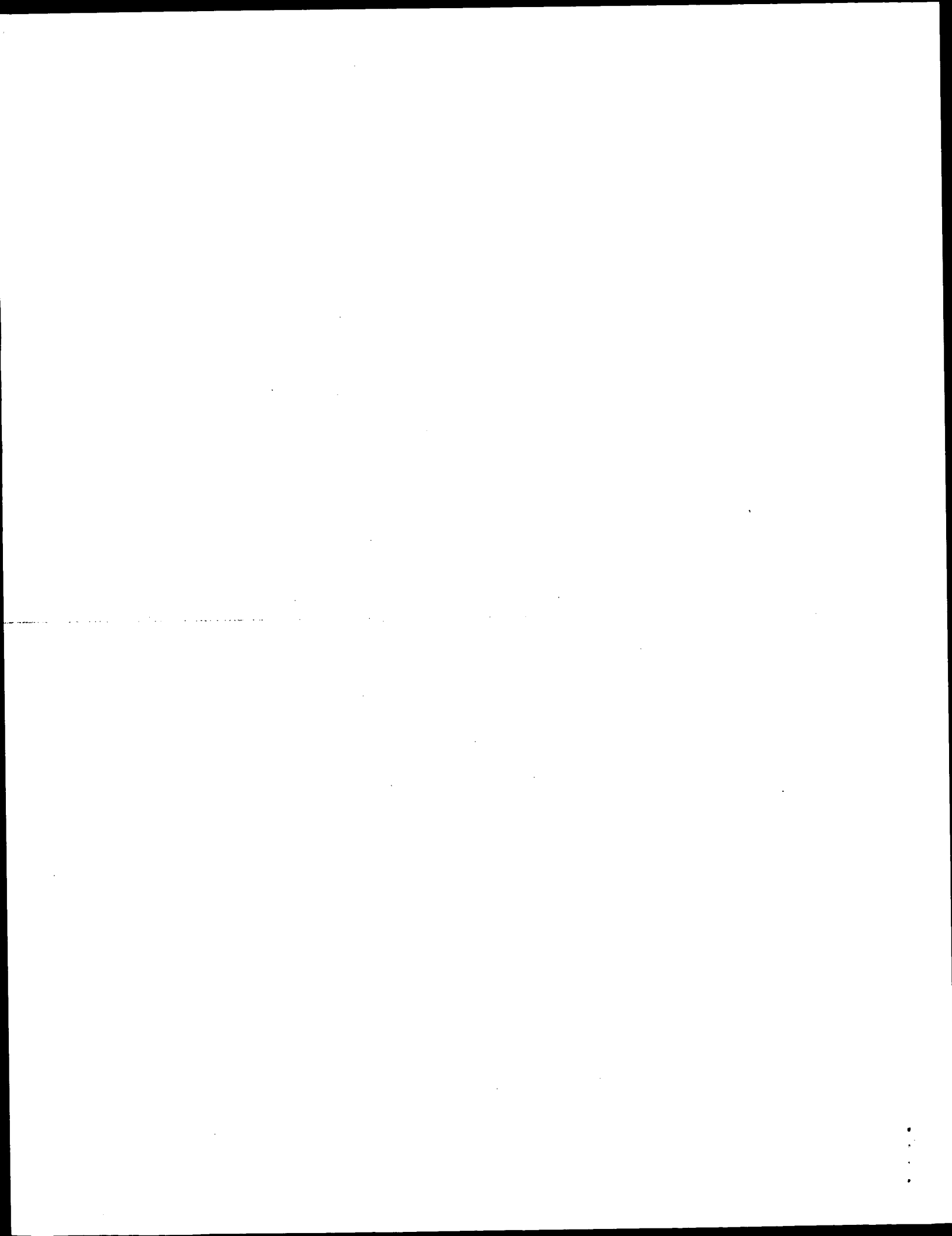
Query Match 5.9%; Score 34.6; DB 19; Length 1105;
 Best Local Similarity 55.3%; Pred. No. 0.32;
 Matches 89; Conservative 0; Mismatches 69; Indels 3; Gaps 1;
 QY 4 gagtatttaattatgcctgttagaattgtttaagaacaggcattcagttcgacaaa 63
 DB 794 gaattatttaactatgcttattagaagaatttaaaatgaattattcaaacgaggaaa 853
 QY 64 taagaaccccttagttatcttgggttgcaccatgctgatttgcgaatgct 123

Db 854 tatgaacccggttttgaaaattg---tagaatatgtggaagtcacatgtgtaatggctaacttc 910

QY 124 ctgcttgattattttgaaaatcagcaggggacagaataaaaaact 164
|| || | | | | | | | | | | | | | | | |

Db 911 cttccgaagggaaggaatcttttcagaaagatcgaataaaagct 951
|| || | | | | | | | | | | | | | | | |

Search completed: June 19, 2002, 16:51:12
Job time: 17530 sec



Qy 241 aatctctgccagccgaagcctgaggtgctaccgccgaagccggaacaata 300
 Db 1039 aatctctgccagccgaagcctgaggtgctaccgccgaagccggaacaata 1098
 Qy 301 tgaatcgagttatggcgccgacacataaaagattccatttggataaagacgaatctgt 360
 Db 1099 tgaatcgagttatggcgccgacacataaaagattccatttggataaagacgaatctgt 1158
 Qy 361 tactttgctgagtcgacgcaacacatttccacgtccaccaatccagtgagtagctgc 420
 Db 1159 tactttgctgagtcgacgcaacacatttccacgtccaccaatccagtgagtagctgc 1218
 Qy 421 tggagaagaaagccgcaagcctctatctacgtccatttgttttgcgcatcggaacgc 480
 Db 1219 tggagaagaaagccgcaagcctctatctacgtccatttgttttgcgcatcggaacgc 1278
 Qy 481 tgagttgatctgg 493
 Db 1279 tgagttgatctgg 1291

RESULT 2

US-08-418-071-3
 ; Sequence 3, Application US/08418071
 ; Patent No. 5846705
 ; GENERAL INFORMATION:
 ; APPLICANT: Wu, Rey-Yuh
 ; APPLICANT: You, Li-Ru
 ; APPLICANT: Soong, Tai-Seng
 ; TITLE OF INVENTION: NUCLEOTIDE SEQUENCE OF TWO CIRCULAR ssDNA
 ; TITLE OF INVENTION: ASSOCIATED WITH BANANA BUNCHY TOP VIRUS AND METHOD FOR DETECT
 ; TITLE OF INVENTION: BANANA BUNCHY TOP VIRUS
 ; NUMBER OF SEQUENCES: 18
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Fish & Neave
 ; STREET: 1251 Avenue of the Americas
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: United States
 ; ZIP: 10020

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Word Perfect 5.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/418,071
 FILING DATE: 06-APR-1995
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Haley Jr., James F.
 REGISTRATION NUMBER: 27,794
 REFERENCE/DOCKET NUMBER: DCB-1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 596-9000
 TELEFAX: (212) 596-9090
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1106 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: circular
 MOLECULE TYPE: genomic DNA
 DESCRIPTION: desc= "BBTV DNA I (clone 7-4-2)"
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 ORIGINAL SOURCE:
 ORGANISM: Banana Bunchy Top Virus (BBTV)

US-08-418-071-3
 Query Match 9.9%; Score 58; DB 2; Length 1106;
 Best Local Similarity 57.9%; Pred. No. 1.4e-10;
 Matches 112; Conservative 0; Mismatches 78; Indels 3; Gaps 1;
 Qy 1 ttacagattattaaattatgacctgttagaattgtttaagaacaggcatttcagttcgac 60
 Db 744 TCAGATTATCTGAATTATGCGCTTATAGACAAATTAAGATAGAGTTTAAATAATACA 803

Matches 103; Conservative 0; Mismatches 75; Indels 0; Gaps 0;
 Qy 4 gagtattaaattatgacctgttagaattgtttaagaacaggcatttcagttcgacaaa 63
 Db 749 GATTATTAAATTATGATCTAATACAGAGCATTAAGGATAGGTTTATAGAGAGTACTAAA 808
 Qy 64 tacgaaccccttagttatcttgggttcgacctgctgcatgtactctgtatttgcgaatctc 123
 Db 809 TACAACCCCAATAAGATAGTTGAATTAGTAAATACATGTAATCGTCATGCGCAATTTC 868
 Qy 124 ctgctgattattgaaatcagcagcgagcagataaaactgtggaattatttaaagta 181
 Db 869 ATGCGCTGACTTCTGTAAATATCTCCGAAGATCGAATAAATAATCATTATTTCGTGAAGAA 926

RESULT 3

US-08-418-071-5
 ; Sequence 5, Application US/08418071
 ; Patent No. 5846705
 ; GENERAL INFORMATION:
 ; APPLICANT: Wu, Rey-Yuh
 ; APPLICANT: You, Li-Ru
 ; APPLICANT: Soong, Tai-Seng
 ; TITLE OF INVENTION: NUCLEOTIDE SEQUENCE OF TWO CIRCULAR ssDNA
 ; TITLE OF INVENTION: ASSOCIATED WITH BANANA BUNCHY TOP VIRUS AND METHOD FOR DETE
 ; TITLE OF INVENTION: BANANA BUNCHY TOP VIRUS
 ; NUMBER OF SEQUENCES: 18
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Fish & Neave
 ; STREET: 1251 Avenue of the Americas
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: United States
 ; ZIP: 10020

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Word Perfect 5.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/418,071
 FILING DATE: 06-APR-1995
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Haley Jr., James F.
 REGISTRATION NUMBER: 27,794
 REFERENCE/DOCKET NUMBER: DCB-1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 596-9000
 TELEFAX: (212) 596-9090
 INFORMATION FOR SEQ ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1091 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: circular
 MOLECULE TYPE: subgenomic DNA
 DESCRIPTION: /desc "BBTV DNA II (clone 2)"
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 ORIGINAL SOURCE:
 ORGANISM: Banana Bunchy Top Virus (BBTV)

US-08-418-071-5
 Query Match 9.0%; Score 52.2; DB 2; Length 1091;
 Best Local Similarity 58.0%; Pred. No. 1.6e-08;
 Matches 112; Conservative 0; Mismatches 78; Indels 3; Gaps 1;
 Qy 1 ttacagattattaaattatgacctgttagaattgtttaagaacaggcatttcagttcgac 60
 Db 744 TCAGATTATCTGAATTATGCGCTTATAGACAAATTAAGATAGAGTTTAAATAATACA 803

QY 61 aaatcagaacc---ccttagttatcttgggttcgaccatgtgcattgactcgtattggc 117
|||||
Db 804 AAATAGCAACCATGCTGATTAGAAAAGATGACAAAATGTCATGTAATTGTTATGGCA 863
QY 118 aatgctcctgcctgattattgaaatcagcaggacagaaataaaactgtggaattttaa 177
|||||
Db 864 AATGCTGCTGATTATTGTAATAATTCAGAGAAGATAGATAAAAATAATTAATTGTTGA 923
QY 178 agtatgtgtcatc 190
| | | | |
Db 924 GAAAGGAAACTTC 936

RESULT 4
US-08-418-071-4
; Sequence 4, Application US/08418071
; Patent No. 5846705
; GENERAL INFORMATION:
; APPLICANT: Wu, Rey-Yuh
; APPLICANT: You, Li-Ru
; APPLICANT: Soong, Tai-Seng
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCE OF TWO CIRCULAR ssDNA
; TITLE OF INVENTION: ASSOCIATED WITH BANANA BUNCHY TOP VIRUS AND METHOD FOR DETECTI
; TITLE OF INVENTION: BANANA BUNCHY TOP VIRUS
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Neave
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 5.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/418,071
; FILING DATE: 06-APR-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley Jr., James F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: DCB-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 596-9000
; TELEFAX: (212) 596-9090
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1096 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular
; MOLECULE TYPE: genomic DNA
; DESCRIPTION: /desc "BBTV DNA II (clone 2-17)"
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Banana Bunchy Top Virus (BBTV)
US-08-418-071-4

Query Match 9.0%; Score 52.2; DB 2; Length 1096;
Best Local Similarity 58.0%; Pred. No. 1.6e-08;
Matches 112; Conservative 0; Mismatches 78; Indels 3; Gaps 1;
QY 1 tttagattattaaatattgcccctgttagaattgtttaagaacagggcattcagttcggac 60
| | | | |
Db 749 TCAGATTATCTGAATTATGCGCTTATAGACAAATTAAGAATAGAGTTTAAATAAATACA 808
QY 61 aaatcagaacc---ccttagttatcttgggttcgaccatgtgcattgactcgtattggc 117
|||||

Db 809 AAATAGCAACCATGCTGATTAGAAAAGATGACAAAATGTCATGTAATTGTTATGGCA 868
QY 118 aatgctcctgcctgattattgaaatcagcaggacagaaataaaactgtggaattttaa 177
|||||
Db 869 AATGCTGCTGATTATTGTAATAATTCAGAGAAGATAGATAAAAATAATTAATTGTTGA 928
QY 178 agtatgtgtcatc 190
| | | | |
Db 929 GAAAGGAAACTTC 941

RESULT 5
US-08-793-634B-6
; Sequence 6, Application US/08793634B
; Patent No. 6211431
; GENERAL INFORMATION:
; APPLICANT: Boevink, Petra C.
; APPLICANT: Surin, Brian P.
; APPLICANT: Keese, Paul K.
; APPLICANT: Chu, Paul W.G.
; APPLICANT: Waterhouse, Peter M.
; APPLICANT: Khan, Rafiqul I.
; APPLICANT: Larkin, Phillip J.
; APPLICANT: Taylor, William C.
; APPLICANT: Marchall, Jerry S.
; TITLE OF INVENTION: NOVEL PLANT PROMOTERS AND USES
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Scully, Scott, Murphy & Presser
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/793,634B
; FILING DATE: June 9, 1997
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Digiglio, Frank S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 10530
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 516-742-4343
; TELEFAX: 516-742-4366
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1017 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-793-634B-6

Query Match 7.8%; Score 45.6; DB 4; Length 1017;
Best Local Similarity 56.4%; Pred. No. 3.3e-06;
Matches 106; Conservative 0; Mismatches 79; Indels 3; Gaps 1;
QY 1 tttagattattaaatattgcccctgttagaattgtttaagaacagggcattcagttcggac 60
| | | | |
Db 726 TCAGATTATCTGAATTATGCTGTAATAAGACAGGTTAAGAAATAGGTAATGTTGAACTACT 785
QY 61 aaatcagaacc---ccttagttatcttgggttcgaccatgtgcattgactcgtattggc 117
|||||
Db 786 AAGTATGAGCCATGCTGTAATGCGGGATGATTAATCATCTCTTCATGTAATTGTTGTTGCA 845
QY 118 aatgtcctgctgattattgaaaatcagcaggacagaaataaaactgtggaattttaa 177
|||||

Db 846 AATGACTCCAGATTTGGGAAATAAAGTCAAGATAGATAAAATTAATTCCTGTTGTA 905
Qy 178 agtatgtg 185
Db 906 AAACCTCG 913

RESULT 6
US-08-202-186-16
; Sequence 16, Application US/08202186
; Patent No. 5756708
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/202,186
; FILING DATE: 24-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: JEFFERY, Donald D.
; REGISTRATION NUMBER: 19,980
; REFERENCE/DOCKET NUMBER: 71611/102 FIKE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1103 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular
US-08-202-186-16

Query Match 5.9%; Score 34.6; DB 1; Length 1103;
Best Local Similarity 55.3%; Pred. No. 0.028;
Matches 89; Conservative 0; Mismatches 69; Indels 3; Gaps 1;
Qy 4 gagtattaaattatgcctttagaattgtttaagaacaggcattcgatcgacaaa 63
Db 792 GAATATTTAAACTAGCTTTATTAGAAGATTTAAAAATGGAATTATTCAAAGCGGAAA 851
Qy 64 tagaacccttagttatcttgggttcgaccatgtgcattgctactgtatttgcgaatg 123
Db 852 TATGAACCCGTTTGGAAAATG---TAGAATATGTGGAGTCATTGTAATGGCTAACTTC 908
Qy 124 ctgcctgattatttgaataatcagcaggacagataaaact 164
Db 909 CTTCGAGGAGGAAGAACTTTTCTGAGATCGAATAAGCT 949

RESULT 7
US-08-202-186-18
; Sequence 18, Application US/08202186
; Patent No. 5756708
; GENERAL INFORMATION:

; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/202,186
; FILING DATE: 24-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: JEFFERY, Donald D.
; REGISTRATION NUMBER: 19,980
; REFERENCE/DOCKET NUMBER: 71611/102 FIKE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1105 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular
US-08-202-186-18

Query Match 5.9%; Score 34.6; DB 1; Length 1105;
Best Local Similarity 55.3%; Pred. No. 0.028;
Matches 89; Conservative 0; Mismatches 69; Indels 3; Gaps 1;
Qy 4 gagtattaaattatgcctttagaattgtttaagaacaggcattcgatcgacaaa 63
Db 794 GAATATTTAAACTAGCTTTATTAGAAGATTTAAAGATTTAAGATTTATCAAAAGCGGAAA 853
Qy 64 tagaacccttagttatcttgggttcgaccatgtgcattgctactgtatttgcgaatg 123
Db 854 TATGAACCCGTTTGGAAAATG---TAGAATATGTGGAGTCATTGTAATGGCTAACTTC 910
Qy 124 ctgcctgattatttgaataatcagcaggacagataaaact 164
Db 911 CTTCGAGGAGGAGGAATCTTTTCAGAGATCGAATAAGCT 951

RESULT 8
US-08-202-186-14
; Sequence 14, Application US/08202186
; Patent No. 5756708
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/202,186
FILING DATE: 24-FEB-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: JEFFERY, Donald D.
REGISTRATION NUMBER: 19,980
REFERENCE/DOCKET NUMBER: 71611/102 FIKE
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202 672 5300
TELEFAX: 202 672 5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 1110 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: circular
US-08-202-186-14

Query Match 5.8%; Score 34; DB 1; Length 1110;
Best Local Similarity 54.9%; Pred. No. 0.047;
Matches 89; Conservative 0; Mismatches 70; Indels 3; Gaps 1;

QY 4 gagtatttaattgcccctgttagaattgtttaagacaggggcattcagttcggaacaa 63
DB 791 GATTATTAAATTTATGGTTTATTAGAGGAATTTAAGAATGGAATAATTCAAAGCGGAAA 850
QY 64 tacaaaccccttagttatcttgggttcgaccatgtgcattcgtattgccaatgctc 123
DB 851 TATGAACCCGTTTGAAGATTG---TAGAATATGTCGAAGTCATTGTAATGGCTAACTTC 907
QY 124 ctgcctgattattgaaatacagcagggacagataaaactg 165
DB 908 CTTCGGAAGGAAGGAATCTTTCTGAGATCGAATAAAGTTG 949

RESULT 9
US-08-202-186-12
Sequence 12, Application US/08202186
Patent No. 5756708
GENERAL INFORMATION:
APPLICANT: KARAN, Mirko
APPLICANT: BURNS, Thomas M.
APPLICANT: DALE, James L.
APPLICANT: HARDING, Robert M.
TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
NUMBER OF SEQUENCES: 60
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W.
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/202,186
FILING DATE: 24-FEB-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: JEFFERY, Donald D.
REGISTRATION NUMBER: 19,980
REFERENCE/DOCKET NUMBER: 71611/102 FIKE
TELECOMMUNICATION INFORMATION:

TELEPHONE: 202 672 5300
TELEFAX: 202 672 5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 1111 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: circular
US-08-202-186-12

Query Match 5.8%; Score 34; DB 1; Length 1111;
Best Local Similarity 54.9%; Pred. No. 0.047;
Matches 89; Conservative 0; Mismatches 70; Indels 3; Gaps 1;

QY 4 gagtatttaattgcccctgttagaattgtttaagacaggggcattcagttcggaacaa 63
DB 792 GATTATTAAATTTATGGTTTATTAGAGGAATTTAAGAATGGAATAATTCAAAGCGGAAA 851
QY 64 tacgaaccccttagttatcttgggttcgaccatgtgcattcgtattgccaatgctc 123
DB 852 TATGAACCCGTTTGAAGATTG---TAGAATATGTCGAAGTCATTGTAATGGCTAACTTC 908
QY 124 ctgcctgattattgaaatacagcagggacagataaaactg 165
DB 909 CTTCGGAAGGAAGGAATCTTTCTGAGATCGAATAAAGTTG 950

RESULT 10
US-08-202-186-17
Sequence 17, Application US/08202186
Patent No. 5756708
GENERAL INFORMATION:
APPLICANT: KARAN, Mirko
APPLICANT: BURNS, Thomas M.
APPLICANT: DALE, James L.
APPLICANT: HARDING, Robert M.
TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
NUMBER OF SEQUENCES: 60
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W.
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/202,186
FILING DATE: 24-FEB-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: JEFFERY, Donald D.
REGISTRATION NUMBER: 19,980
REFERENCE/DOCKET NUMBER: 71611/102 FIKE
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202 672 5300
TELEFAX: 202 672 5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 1104 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: circular
US-08-202-186-17

Query Match 5.7%; Score 33; DB 1; Length 1104;

NAME: JEFFERY, Donald D.
REGISTRATION NUMBER: 19,980
REFERENCE/DOCKET NUMBER: 71611/102 FIVE
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202 672 5300
TELEFAX: 202 672 5399
TELEX: 904136

[illegible]

```

14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046

```

```

;
; NAME/KEY: CDS
; LOCATION: 129..989
; FEATURE:
US-08-202-186-24

```

	Query Match	5.6%	Score 32.4;	DB 1;	Length 1111;
	Best Local Similarity	54.3%;	Pred. No. 0.17;		
	Matches 88;	Conservative 0;	Mismatches 71;	Indels 3;	Gaps 1;
Qy	4	qagttattaaattatgccctgttagaatgtttaagacacaggcgatcagttcgcgacaaa	63		
Db	819	GAITATTATTAATTATGGCTTATTAGAGGAATTTAAGAAATGGAATAATTCAAAGCGGGNAA	878		
Qy	64	tcagaaaccccttagttatcttgggttcgaccatgtgatctcgtatttgcgcaatgtc	123		
Db	879	TATGAACCCGTTTTGAAGATAG---TAGAATATGTCGAAGTCATTGTAATGGTGTAAC TTC	935		
Qy	124	ctgcctgattatttgaataatcagcagggacagataaaactg	165		
Db	936	CTTCCGAAGGAAGAAATCTTTTCTGAAGATCGAATAAAGTTG	977		

```

RESULT 15
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

```

```

Query Match      5.5%; Score 32.2; DB 4; Length 4411529;
Best Local Similarity 57.4%; Pred. No. 22;
Matches 59; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

Qy 410 gagttagttgctgagagaagacgcgaagcctctctatctaccgtccattgtttttggcg 469
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3577728 GAGCGGCACGGCGGAACAAACACCGCGAGCTGCCGTCTCCGTCGGATCGGTGTTGCAC 35777659

Qy 470 gatcgagcgctgagttgatctggtgcgcacaaaacctctgct 510
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 35777668 GCACCTGTTTCCGAGCGCGGCACGAGCGGAATCCCAACTGCT 35777628

```

Search completed: June 19, 2002, 16:09:16
Job time: 15014 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2002, 14:04:20 ; Search time 7489.97 Seconds
(without alignments)
1050.568 Million cell updates/sec

Title: US-09-462-955B-1_COPY_409_991

Perfect score: 583

Sequence: 1 tttagagtatttaattatgc.....ccccaacctctgtaacccc 583

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:**
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
C 1	40	6.9	512	12	AQ121628 HS_3089_A
2	38	6.5	2275	9	AF034173 AF034173
3	36.8	6.3	498	12	AQ615572 HS_5144_B
4	36.4	6.2	787	12	CNS010B7 AL098749 Drosophil
C 5	36	6.2	507	12	A2444169 A2444169 IM02339F20
C 6	36	6.2	929	12	BH137680 BH137680 ENTPC13TR
C 7	35.6	6.1	318	12	AQ089903 HS_3001_A
8	35.6	6.1	1101	12	CNS016H0 AL106734 Drosophil
9	35.4	6.1	460	9	AI942816 fc67b05_x
10	35.4	6.1	488	12	CNS00Q0X AL085215 Arabidops
11	35.2	6.0	848	12	AQ746046 HS_2277_A
12	35	6.0	524	12	BH175883 BH175883 004_P_17-
13	35	6.0	524	12	CNS07IRP AL612823 T3 end of
14	34.6	5.9	487	12	AQ418525 RPCI-11-2
15	34.6	5.9	555	12	AQ417598 RPCI-11-2
16	34.6	5.9	665	9	BB633121 BB633121
C 17	34.6	5.9	689	12	AG167308 Pan trogl

C 18	34.4	5.9	594	12	AZ397924
19	34.4	5.9	898	9	AL667528
C 20	34.2	5.9	345	12	AZ401897
C 21	34	5.8	442	12	AQ228962
C 22	34	5.8	443	12	AQ010299
C 23	34	5.8	1101	12	CNS00DBS
C 24	33.8	5.8	410	12	AQ807176
C 25	33.8	5.8	673	12	AG167164
C 26	33.8	5.8	749	12	AZ331069
C 27	33.8	5.8	772	12	AQ488982
C 28	33.8	5.8	873	12	CNS012PA
C 29	33.8	5.8	1174	10	BE779656
C 30	33.6	5.8	526	12	AQ008775
C 31	33.6	5.8	403	10	BG006561
C 32	33.6	5.8	418	12	AQ632834
C 33	33.6	5.8	434	12	AQ728714
C 34	33.6	5.8	605	12	AQ481591
C 35	33.6	5.8	658	12	AQ077737
36	33.4	5.7	283	9	AI065300
C 37	33.4	5.7	348	12	AQ348508
C 38	33.4	5.7	411	9	AI226907
C 39	33.4	5.7	460	12	AQ624335
C 40	33.4	5.7	517	12	AQ565229
C 41	33.4	5.7	532	12	AQ565203
C 42	33.4	5.7	637	12	AQ377093
C 43	33.4	5.7	751	12	AQ745275
C 44	33.4	5.7	764	12	AQ240040
C 45	33.4	5.7	785	11	AK002866

ALIGNMENTS

RESULT 1
AQ121628/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

FEATURES

source

AQ121628 512 bp DNA linear GSS 22-SEP-1998
HS_3089_AL_B03_MF CIT Approved Human Genomic Sperm Library D Homo
sapiens genomic clone Plate=3089 Col=5 Row=C, DNA sequence.
AQ121628
AQ121628.1 GI:3498794
GSS -
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 512)
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 3089 row: C column: 5
Class: BAC ends
High quality sequence stop: 512.
Location/Qualifiers
1..512
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="plate=3089 Col=5 Row=C"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in
E-Coli DH10B"


```

KEYWORDS      GSS.
SOURCE        fruit fly.
ORGANISM      Drosophila melanogaster
REFERENCE     Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
AUTHORS       Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
TITLE         Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
JOURNAL       1 (bases 1 to 787)
              Direct Submission
              Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
              BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
              - Web : www.genoscope.cns.fr)
COMMENT       Determination of this BAC-end sequence was carried out as part of a
              collaboration with the European Drosophila Genome project (EDGP) -
              http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
              library (Dros BAC) was made by Alain Billaud at CEPH (Centre
              d'Etude du Polymorphisme Humain) with funding provided by a MRC
              project grant. The DNA was prepared from embryos by Alain Bucheton
              and Genevieve Payan. It has been constructed in the vector
              pBeloBAC11.
FEATURES      Location/Qualifiers
              1..787
              /organism="Drosophila melanogaster"
              /plasmid="pBeloBAC11"
              /db_xref="taxon:7227"
              /clone_lib="DrosBAC"
              /clone="BACN03M07"
              /note="end : 77"
BASE COUNT   52 a 114 c 73 g 19 t 529 others
ORIGIN
Query Match      6.2%; Score 36.4; DB 12; Length 787;
Best Local Similarity 3.0%; Pred. No. 2.2;
Matches 10; Conservative 109; Mismatches 215; Indels 0; Gaps 0;
QY 189 tctaaattacaccataccgccgccacgcgtatgtttacattcttatgaatactct 248
Db 425 KNTMMNNKNCMMNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNN 484
QY 249 gccagcgccgaagcctggagtgctaccgccgaagccggaagccggaacaataatga 308
Db 485 KNNMKKMMKMMNTNNMKNTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 544
QY 309 gttatggcgccgcacataaagattccatttggaataagaacgaatctgttacttgc 368
Db 545 NMNTNNNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNN 604
QY 369 ttgcagtcgacgcgaaccattccacgtccacgaatccaggtgagtagctgtggagaag 428
Db 605 NNMKNMKTMMNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 664
QY 429 aaagccgcaagcctctatctaccgtccattgttttgcgcgagtcggcgcgtgagttga 488
Db 665 TNNNNNNNNKNTMMNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 724
QY 489 tctggcgcaaaacctctgtgaagtcctgcctgct 522
Db 725 MNKTNKNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 758
RESULT 5
AZ444169/c 507 bp DNA linear GSS 04-OCT-2000
LOCUS      1M0239P20F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
DEFINITION AZ444169
ACCESSION  AZ444169.1 GI:10592981
VERSION     GSS.
KEYWORDS    house mouse.
SOURCE      Mus musculus
ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE   1 (bases 1 to 507)
AUTHORS     Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D., Weiss,R.
            Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
            Unpublished (2000)
            Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000 Std Error: 0.00
            Plate: 0239 row: F column: 20
            Seq primer: CGTTGTAAACGACGCGCCAGT
            Class: plasmid ends
            High quality sequence stop: 507.
FEATURES      Location/Qualifiers
              1..507
              /organism="Mus musculus"
              /strain="C57BL/6J"
              /db_xref="taxon:10090"
              /clone="UUGC1M0239F20"
              /clone_lib="Mouse 10kb plasmid UUGC1M library"
              /sex="Male"
              /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
              /note="Vector: PWD42nv; Purified genomic DNA from M.
              musculus C57BL/6J (male) was obtained from the Jackson
              Laboratory Mouse DNA Resource
              (http://www.jax.org/resources/documents/dnares/). The DNA
              was hydrodynamically sheared by repeated passage through a
              0.005 inch orifice at constant velocity. The sheared DNA
              was blunt end-repaired with T4 DNA polymerase and T4
              polynucleotide kinase. Adaptor oligonucleotides were
              ligated to the blunt ends in high molar excess. The
              adaptor DNA was purified and size-selected for a 9.5 to
              10.5 kb range using preparative agarose gel
              electrophoresis. Vector DNA was prepared from a derivative
              of PWD42 (gil4732114|gb|AF129072.1), a copy-number
              inducible derivative of plasmid p1. The vector was ligated
              with adaptors complementary to the insert adaptors and
              purified. The sheared, adaptor mouse DNA was annealed to
              adaptor vector DNA, and transformed into
              chemically-competent E. coli XL10-Gold (Stratagene) cells
              and selected for ampicillin resistance."
BASE COUNT   205 a 74 c 92 g 136 t
ORIGIN
Query Match      6.2%; Score 36; DB 12; Length 507;
Best Local Similarity 58.3%; Pred. No. 2.4;
Matches 63; Conservative 0; Mismatches 45; Indels 0; Gaps 0;
QY 96 tgtgcagtactcgtatttgcgaatgctcctgcctgatttattgaaatccgaggagacag 155
Db 364 TATTTCATGAACAGAAATGGGAAGCATTCTCTATTAATTAATGATCAGTACGCCATAT 305
QY 156 aataaaactgtggaatttttaagatgtgtcattcattaaattacacaa 203
Db 304 GATATTGCTCTGACAGGTTCAATCATCTTTTATTTTAAATAACAACAA 257
RESULT 6
BH137680/c 929 bp DNA linear GSS 07-AUG-2001
LOCUS      ENTPL137R Entamoeba histolytica Sheared DNA Entamoeba histolytica
DEFINITION genomic, DNA sequence.
ACCESSION  BH137680
VERSION     BH137680.1 GI:15096741

```

GSS.
Entamoeba histolytica.
SOURCE Entamoeba histolytica
ORGANISM Eukaryota; Entamoebidae; Entamoeba.
REFERENCE 1 (bases 1 to 929)
AUTHORS Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.
TITLE Determination of clone end sequences from Entamoeba histolytica
HMI:IMSS sheared DNA library (2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjloftus@ig.rockefeller.edu
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
DNA library
Seq primer: M13-Reverse
Class: shotgun
High quality sequence start: 17
High quality sequence stop: 683.
Location/Qualifiers

```

1. 323 /organism="Entamoeba histolytica"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica sheared DNA"
/notes="vector: pHO1; Site:1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + i method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaubin and B.
Barell, Oxford University Press, 1999)."
288 a 114 c 135 q 392 t

```

RESULT	7
AQ089903/c	
LOCUS	
DEFINITION	318 bp DNA linear GSS 26-AUG-1998
ACCESSION	AQ089903
VERSION	AQ089903.1 GI:3458814
DESCRIPTION	HS-3001_A2_H04_MR CTT Approved Human Sperm Library D Homo sapiens genomic clone Plate=3001 Col=8 Row=0, DNA sequence.

KEYWORDS	GSS.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Muthera; Primates; Catarrhini; Hominoidea; Homo.
AUTHORS	1 (bases 1 to 318) Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE	Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL	Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE	99380589
COMMENT	Contact: Mahairas GG, Wallace JC, Hood L High Throughput Sequencing Center University of Washington 401 Queen Anne Avenue North, Seattle, WA 98109, USA Tel: (206) 616-3618 Fax: (206) 616-3887 Email: jwallace@u.washington.edu Sequence Tagged Connector Plate: 3001 row: 0 column: 8 Class: BAC ends

	Query Match	6.1%;	Score 35.6;	DB 12;	Length 318;
	Best Local Similarity	54.6%;	Pred. No. 2.6;		
	Matches	71;	Conservative	0;	Mismatches
					Indels
					0;
					Gaps
Qy	34	gttaagaacaggcgatcagttcggacacaaatcacgaaccccttagttatctctgggtcgac	93		
Db	212	gttctggccaggcgaaatcaggcagcaggaagaaataaagggtatttaattaggaaagagag	153		
Qy	94	catgtgcagtctactcgctattttgccaatgtcctgcctgattatttgaaatcagcaggagc	153		
Db	152	gatttcaaatgtccctctttgcagatgcacatgactgtgtatctcagaaaaacggcatgttc	93		
Qy	154	agaataaac	163		
Db	92	tcagttccaac	83		

RESULT	8
CNS016H0	
LOCUS	linear
DEFINITION	1101 bp DNA GSS 26-JUL-1999 Drosophila melanogaster genome survey sequence T7 end of BAC BACN16B20 of DrosBAC library from Drosophila melanogaster (fruit fly) genomic survey sequence.
ACCESSION	AL106734
VERSION	AL106734.1 GI:5623558
KEYWORDS	GSS.
SOURCE	fruit fly.
ORGANISM	Drosophila melanogaster
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE	1 (bases 1 to 1101)

REFERENCE
AUTHORS
TITLE
JOURNAL
Genoscope.
Direct Submission
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : secref@genoscope.cns.fr)
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1101)
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Seq primer: T7 ET from Amersham
High quality sequence stop: 443.

```

1..460
/organism="Danio rerio"
/db_xref="taxon:7955"
/clone="IMAGE:3726417"
/clone_lib="Zebrafish WashU MPIMG EST"
/sex="mixed"
/tissue_type="36 somite embryos, adult livers, shield
stage embryos"
/lab_host="XLI-blue MRF"
/note="vector: pSPORT1; Site_1: NotI; Site_2: SalI; 1st
strand cDNA was primed with a Not I - oligo(dNTP)15 primer
[5',GCACTAGTCTCTAGATCGCAGGCCGCCCTTTTTTTTTTTT3'];

```

```

/lab_host="XLI-blue MRF"
/notes="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; 1st
strand cDNA was primed with a Not I - oligo(dT)15 primer
[5'-GACATAGATTCTAGATCGGAGCGGCGCCCTTTTTTTTTTTT3-];
double-stranded cDNA

```

Institut fuer Molekulare Genetik, Berlin). cDNAs for EST analysis were selected following oligonucleotide hybridization fingerprinting of arrayed clones from zebrafish late somitogenesis (26 ss), adult liver or embryonic shield stage (5.6 h) libraries. Fingerprint data were used to computationally cluster cDNAs, and a single cDNA from each cluster was chosen for sequencing. In some cases multiple members of the same cluster were

154 a 60 c 59 g 187 t 6.1% Score 35.4; DB 9; Length 460;

Similarity 50.9%; Pred. No. 3.6;
84; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

CCTCCATAATTAATGTTGTTTAAGCTCTATTAAATCATGATGTTGAACAACACAGCAG 124
 gattatttggaataatcagcagggacagataaaactgtggaat 171
 TAATTTTGTGAATTCATCTAGTGAAATTAATATTCATGTCGCAT 169

NS00Q0X 488 bp DNA linear GSS 28-JUN-1999
Arabidopsis thaliana genome survey sequence SP6 end of BAC F10F11
ICF library from strain Columbia of Arabidopsis thaliana,
genomic survey sequence.
AL085215
AL085215.1 GI:5286355

SS.
thal. cress.
Arabidopsis
Arabidopsis thaliana
Cucurbitaceae; viridiplantae;
Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
eurosids II; Brassicales; Brassicaceae; Arabidopsi-
(bases 1 to 488)
Salmonella M...
N...
P...

Samson, D., Saurin, W., Weissenbach, J. and Quétier, F.

FEATURES	source	1. 848	Location/Qualifiers
		/organism="Homo sapiens"	
		/db_xref="taxon:9606"	
		/clone="Plate-2277 Col-22 Row-K"	
		/clone_lib="CIT Approved Human Genomic Sperm Library D"	
		/sex="male"	
		/note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in E-Coli DH10B"	
BASE COUNT		228 a 136 c 117 g 144 t	223 others
ORIGIN			
	Query Match	6.0%;	Score 35.2; DB 12; Length 848;
	Best Local Similarity	55.8%;	Pred. No. 5.6;
	Matches 67;	Conservative	0; Mismatches 53; Indels 0; Gaps 0;
QY	34	gttaagaacagggcattcgattggcgacaatacgaaccccttagttatcttgggttcgac	93
Db	334	GTCTGCCAGGCGCAATCAGGTAGGAGAAAGAAATAACCGGTATTCAATTAGGAAAAGAG	393
QY	94	catgtcagtactcgtatttgcgaatgtcctgcctgatttgaataatcacaggagac	153
Db	394	GAAGTCAAAATGTCCCTGTTTGCAGATGCATGATGTATATTAGAAAACCCCATGTC	453
RESULT	12		
BH175883		524 bp	DNA linear
LOCUS	004_P_17-rev SmbAC1	Schistosoma mansoni	genomic clone 004P17 5',
DEFINITION		DNA sequence.	
ACCESSION	BH175883		
VERSION	BH175883.1	GI:16274497	
KEYWORDS	GSS.		
SOURCE	Schistosoma mansoni.		
ORGANISM	Schistosoma mansoni		
REFERENCE	Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;		
AUTHORS	Strigeliida; Schistosomatoidea; Schistosomatidae; Schistosoma.		
	1 (bases 1 to 524)		
	Le Paslier,M.-C., Pierce,R.J., Merlin,F., Hirai,H., Wu,W., Williams		
	,D.L., Johnston,D., Loverde,P.T. and Le Paslier,D.		
TITLE	Construction and characterization of a Schistosoma mansoni		
	bacterial artificial chromosome library		
JOURNAL	Genomics 65 (2), 87-94 (2000)		
MEDLINE	20247247		
COMMENT	Other_GSSs: 004_P_17-21		
	Contact: Pierre RJ		
	INSERM U 167		
	Institut Pasteur de Lille		
	1 rue du Professeur A. Calmette, 59019-Lille, France		
	Tel: (33) (0)3 20877783		
	Fax: (33) (0)3 20877888		
	Email: Raymond.Pierce@pasteur-lille.fr		
	CNS sequencing ID=DG0AA004CH0BP1 Bases 1-130 have 86% identity to		
	S.mansoni EST AI974978.1 from base 292-163. Bases 283-408 have 84%		
	identity to S.mansoni EST AI394783.1 from base 448-326.		
	Plate: 004 row: P column: 17		
	Seq primer: M13 reverse primer		
	Class: BAC ends		
	High quality sequence stop: 524.		
FEATURES	Location/Qualifiers		
source	1. 524		
	/organism="Schistosoma mansoni"		
	/strain="Puerto-Rican"		
	/db_xref="taxon:6183"		
	/clone="004P17"		
	/clone_lib="SmbAC1"		
	/sex="mixed"		
	/dev_stage="Cercariae"		
	/lab_host="Biomphalaria glabrata"		
	/note="Vector: pBelOBAC 11; Site_1: Hind III; Partially		
	Hind III digested and size-selected S. mansoni cercarial		
	DNA was ligated into Hind III digested pBelOBAC 11 vector		

Db 59 AGNTATTCGGATCCGTTGGCCAGACACATATCAGCAACAAGCTAGTGTGGGACWAAACA 118

Qy 329 aaagattccattgtgataagaacgaatctgttactttgttcagtcagcgcaaccact 388
| | | | | | | | | | : : | | | | | |

Db 119 ACCANATTCCAGTTGGAGAGAAAATCAACAGANNRYRYTCTGGAGCTGATANGACAC 178

ranges from 70-127 kb and genome coverage is 7.9-fold.			
	101 a	101 c	99 t
153 a	101 c	126 g	99 t
			45 others

Db 179 TTTCMGGAAATCCCCCCA 196

RESULT 14
A0418525

LOCUS	AQ418525	487 bp	DNA	linear	GSS 23-MAR-1999

DEFINITION	RPCI-11-203I8.TV RPCI-11 Homo sapiens genomic clone RPCI-11-203I8, DNA sequence
<p> 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515</p>	

Accession
AQ418525

VERSION AQ418525.1 GI:4476249
KEYWORDS GSS

SOURCE human.

ORGANISM Homo sapiens
Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS	Zhao, S., Adams, M.D., Nierman, W., Malek, J., de Jong, P. and Venter, J.C.
TITLE	Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready Map Building
JOURNAL	Unpublished (1997)
COMMENT	Other GSs. RPCI-11-20318 T1

Contact: Shaying Zhao, William Nierman, Mark Adams

Department of Eukaryotic Genomics
The Institute for Genomic Research

7912 Medical Center Dr., Rockville, MD 20850

Tel: 301 838 0200
Fax: 301 838 0208

Email: hbe@tigr.org

Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong

(pieter@dejong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from Research Genet cs (info@resgen.com) BAC and search page:

http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.

Seq primer: T7
Class: RAC ends

FEATURES

```
source
1. .487
/organism="Homo sapiens"
```

```
/organism Homo sapiens
/db_xref="GDB:757776"
```

```

/db_xref="taxon:9606"
/clone="ppci-11-20318"

```

```
/clone_lib="RPCI-11"
```

```
/sex="Male"  
/cell type="Lymphocytes"
```

```

/cell_type= Lymphocytes
/note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;

```

BASE COUNT	165 a	102 c	96 g	124 +
	RPC111 Human Male BAC Library"			

DATE COUNT	100 g	102 C	30 g	124 C
ORIGIN				

Query Match 5.9%; Score 34.6; DB 12; Length 487;

Best Local Similarity 54.3%; Pred. No. 6.7;
Matches 70. Conservative 0. Mismatches 50. Indels 0. Cans 0.

matches 0, conservative 0, mismatches 33, indels 0, gaps 0;

Qy 20 ccctgtagaatgtgttaagaacagggcattcagttcggacaaatacgaaccccttagtt 79

Db 127 CTCTGGTAGAATTCGGTCCGCTCAGGGCAATCAGGCAGGAGAAAGAAATAAAGGGTGTC 186

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

[illegible]

Db 187 AATTAGGAAAGAGGAAGTCAAATTGTCCCTGTTGCAGATGACATGATTGTGTATCTAC 246

Qy 140 aaatcaqca 148

```

Db      247 AAACCCCA 255

RESULT 15
AQ417598
LOCUS   RP41-11-203C8-TV RPCI-11 Homo sapiens genomic clone RP41-11-203C8,
DEFINITION
DNA sequence.
ACCESSION   AQ417598
VERSION     AQ417598.1 GI:4475441
KEYWORDS    GSS.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 555)
AUTHORS     Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
,J.C.
TITLE       Use of BAC End Sequences from Library RP41-11 for Sequence-Ready
Map Building
JOURNAL     Unpublished (1997)
COMMENT     Other_GSSs: RP41-11-203C8.TJ
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbe@tigr.org
Clones are derived from the human BAC library RP41-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
Research Genet cs (info@resgen.com). BAC end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: T7
Class: BAC ends.
FEATURES             Location/Qualifiers
     source            1..555
                        /organism="Homo sapiens"
                        /db_xref="GDB:757623"
                        /db_xref="taxon:9606"
                        /clone="RP41-11-203C8"
                        /clone_lib="RP41-11"
                        /sex="Male"
                        /cell_type="Lymphocytes"
                        /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
RP4111 Human Male BAC Library"
BASE COUNT          192 a 114 c 110 g 139 t
ORIGIN

Query Match          5.9%; Score 34.6; DB 12; Length 555;
Best Local Similarity 54.3%; Pred. No. 7.1;
Matches 70; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

Qy  20 ccctgttagaattgttaagaacaggcattcagttcggacaaatacgaaccccttagtt 79
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  150 CTCGTGTAAGATTCCGTCGTCAGGGCAATCAGGCAGGAGAAAGAAATAAAGGGGTGTC 209
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy  80 atcttggttcgaccatgtgcattctactcgtatttgccaatgtcctgcctgattattga 139
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  210 AATTAGGAAAAGAGGAAGTCAATTTGTCCTGTTTGCAGATGACATGATTGTGTATCTAC 269
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy  140 aaatcagca 148
    | | | | | | | |
Db  270 AAACCCCA 278

```

Search completed: June 19, 2002, 14:04:23
Job time: 7521 sec

corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated repeat sequence elements. Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at <http://www.sanger.ac.uk/projects/C.elegans/wormpep> This sequence was generated from part of bacterial clone contigs of human chromosome 13, constructed by the Sanger Centre Chromosome 13 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr13>
RP11-29716 is from the library RP11-272L14 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>
VECTOR: pBACe3.6

This sequence is the entire insert of clone RP11-29716 The true left end of clone RP11-207D10 is at 130589 in this sequence. The true right end of clone RP11-272L14 is at 88495 in this sequence.

FEATURES

Source	Location/Qualifiers
misc_feature	1..220115 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="13" /clone="RP11-29716" /clone_lib="RPC1-11.2"
	131..1822 /note="CpG island"
	/evidence=not_experimental 2357..2504 /note="37 copies 4 mer atat 75% conserved"
repeat_region	2363..2502 /note="70 copies 2 mer at 77% conserved"
repeat_region	3453..3755 /note="AluY repeat: matches 1..303 of consensus 5180..5450"
repeat_region	5939..5978 /note="AluSc repeat: matches 1..270 of consensus"
repeat_region	9204..9248 /note="10 copies 4 mer tgga 100% conserved"
repeat_region	10198..10397 /note="15 copies 3 mer gtt 75% conserved"
repeat_region	12023..12404 /note="MER20 repeat: matches 7..218 of consensus"
repeat_region	13891..14196 /note="MSTD repeat: matches 1..394 of consensus"
repeat_region	15743..16019 /note="AluY repeat: matches 1..302 of consensus"
repeat_region	16034..16726 /note="MLTID repeat: matches 190..505 of consensus"
repeat_region	16727..17162 /note="MLT1-INTERNAL repeat: matches 42..880 of consensus"
repeat_region	19077..19259 /note="MLTID repeat: matches 40..482 of consensus"
repeat_region	19751..20050 /note="L2 repeat: matches 2335..2504 of consensus"
repeat_region	20222..20521 /note="AluSg repeat: matches 1..302 of consensus"
repeat_region	25657..25831 /note="AluSg repeat: matches 1..297 of consensus"
repeat_region	26571..26606 /note="Charlie5 repeat: matches 2373..2585 of consensus"
repeat_region	27856..28166 /note="9 copies 4 mer acac 100% conserved"
repeat_region	28333..28578 /note="AluY repeat: matches 1..311 of consensus"
repeat_region	28668..28789 /note="MIR repeat: matches 7..260 of consensus"
repeat_region	/note="AluJb repeat: matches 1..124 of consensus"
repeat_region	29414..29579 /note="MIR repeat: matches 66..259 of consensus"
repeat_region	29712..29855 /note="L2 repeat: matches 1508..1650 of consensus"
repeat_region	30170..30368 /note="L2 repeat: matches 2056..2275 of consensus"
repeat_region	30667..30744 /note="L2 repeat: matches 2672..2749 of consensus"
repeat_region	32012..32087 /note="L1ME2 repeat: matches 6058..6131 of consensus"
misc_feature	32363..33462 /note="CpG island"
repeat_region	/evidence=not_experimental 32558..32703 /note="73 copies 2 mer cc 63% conserved"
misc_feature	32623..32754 /note="Sequence from overlapping clone BA272L14 (ALI38689). Assembly confirmed by restriction digest."
repeat_region	33621..33682 /note="MIR repeat: matches 85..148 of consensus"
repeat_region	33747..33968 /note="MIR repeat: matches 15..257 of consensus"
repeat_region	34151..34515 /note="L1M4 repeat: matches 4316..4702 of consensus"
repeat_region	34520..34746 /note="MLT1A1 repeat: matches 143..365 of consensus"
repeat_region	34756..34890 /note="MLT1A1 repeat: matches 3..138 of consensus"
repeat_region	34891..35118 /note="MLT1A1 repeat: matches 162..365 of consensus"
repeat_region	35316..35546 /note="MLT1A1 repeat: matches 133..365 of consensus"
repeat_region	35547..35633 /note="MLT1A1 repeat: matches 63..138 of consensus"
repeat_region	35900..36016 /note="MLT1A1 repeat: matches 1..117 of consensus"
repeat_region	36404..36604 /note="MLT1C repeat: matches 264..466 of consensus"
repeat_region	36715..37003 /note="AluSg1 repeat: matches 1..289 of consensus"
repeat_region	37056..37085 /note="15 copies 2 mer aa 93% conserved"
repeat_region	37616..37989 /note="THE1C repeat: matches 1..371 of consensus"
repeat_region	39265..39720 /note="MER92B repeat: matches 147..626 of consensus"
repeat_region	39732..40157 /note="THE1C repeat: matches 1..426 of consensus"
repeat_region	40158..40295 /note="MER92B repeat: matches 1..148 of consensus"
repeat_region	40338..42198 /note="L1MCb repeat: matches 115..2032 of consensus"
repeat_region	42660..42698 /note="MER5A repeat: matches 146..187 of consensus"
repeat_region	42996..43434 /note="MLT1C repeat: matches 1..497 of consensus"
repeat_region	44101..44184 /note="MLT1H repeat: matches 448..531 of consensus"
repeat_region	44362..44400 /note="MLT1G repeat: matches 231..269 of consensus"
repeat_region	44516..44571 /note="MLT1H repeat: matches 72..127 of consensus"
repeat_region	45161..45458 /note="AluSg repeat: matches 1..298 of consensus"
repeat_region	45700..45837 /note="MIR repeat: matches 28..236 of consensus"
repeat_region	46028..46270 /note="MER45C repeat: matches 1..228 of consensus"
repeat_region	46466..46695 /note="MER45C repeat: matches 714..952 of consensus"
repeat_region	47322..47737 /note="MSTC repeat: matches 1..405 of consensus"
repeat_region	48351..48562

Quality Coverage: 100%

Total: 100%

100

100

100


```

Query Match      9.1%; Score 34.8; DB 9; Length 40709;
Best Local Similarity 52.8%; Pred. No. 18;
Matches 75; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 72 ctaccggccgaagcggcggaacaatgaatgagttatggcgggccacacataaag 131
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 31651 CTTCCAGTCCCGACGACGTGTGAGAAATAAATCCATGTTGAAGCCAAAGAAAAA 31592
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 132 attccatttgataagaacgaatctgttacttgccttgagtcagtcacgcaacactttcca 191
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 31591 ATTCCATTGTGCTCCTGACTACTCTTTTATTTTATTTTGTGAGACAGAGTCTCAC 31532
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 192 cgtcacaatccaggtgagtag 213
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 31531 CGTGGCCAGGCTGGAGTGCAG 31510
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 7
AC000026/c
LOCUS          AC000026               126312 bp    DNA    linear    PRI 01-JUN-2000
DEFINITION    Homo sapiens Chromosome 22q12 BAC Clone 58b8 In Meningioma Deletion
              Region, complete sequence.
ACCESSION     AC000026
VERSION       AC000026.3  GI:5881547
KEYWORDS      HIG.
SOURCE        human.
ORGANISM      Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1 (bases 1 to 126312)
AUTHORS       Pan,H., Hartman,K., Willingham,D. and Roe,B.A.
TILE         Homo sapiens Chromosome 22q12 BAC Clone 58b8 In Meningioma Deletion
              Region
              Unpublished
REFERENCE     2 (bases 1 to 126312)
AUTHORS       Kim,U.-J. and Simon,M.
JOURNAL       Unpublished
REFERENCE     3 (bases 1 to 126312)
AUTHORS       Peyrard,M. and Dumanski,J.P.
JOURNAL       Unpublished
REFERENCE     4 (bases 1 to 126312)
AUTHORS       Roe,B.A.
TILE         Direct Submission
JOURNAL       Submitted (21-OCT-1996) Department Of Chemistry And Biochemistry,
              The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
              OK 73019, USA
RMARK         58b8
REFERENCE     5 (bases 1 to 126312)
AUTHORS       Roe,B.A.
TILE         Direct Submission
JOURNAL       Submitted (24-FEB-1998) Department Of Chemistry And Biochemistry,
              The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
              OK 73019, USA
REFERENCE     6 (bases 1 to 126312)
AUTHORS       Roe,B.A.
TILE         Direct Submission
JOURNAL       Submitted (03-MAR-1998) Department Of Chemistry And Biochemistry,
              The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
              OK 73019, USA
REFERENCE     7 (bases 1 to 126312)
AUTHORS       Roe,B.A.
TILE         Direct Submission
JOURNAL       Submitted (21-MAR-1998) Department Of Chemistry And Biochemistry,
              The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
              OK 73019, USA
REFERENCE     8 (bases 1 to 126312)
AUTHORS       Roe,B.A.
TILE         Direct Submission
JOURNAL       Submitted (27-MAY-1998) Department Of Chemistry And Biochemistry,
              The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
              OK 73019, USA
REFERENCE     9 (bases 1 to 126312)
AUTHORS       Roe,B.A.

```

```

TITLE          Direct Submission
JOURNAL        Submitted (23-JUN-1998) Department Of Chemistry And Biochemistry,
              The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
              OK 73019, USA
REFERENCE       10 (bases 1 to 126312)
AUTHORS        Roe,B.A.
TILE          Direct Submission
JOURNAL        Submitted (01-AUG-1998) Department Of Chemistry And Biochemistry,
              The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
              OK 73019, USA
REFERENCE       11 (bases 1 to 126312)
AUTHORS        Roe,B.A.
TILE          Direct Submission
JOURNAL        Submitted (01-OCT-1998) Department Of Chemistry And Biochemistry,
              The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
              OK 73019, USA
REFERENCE       12 (bases 1 to 126312)
AUTHORS        Roe,B.A.
TILE          Direct Submission
JOURNAL        Submitted (14-OCT-1998) Department Of Chemistry And Biochemistry,
              The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
              OK 73019, USA
REFERENCE       13 (bases 1 to 126312)
AUTHORS        Roe,B.A.
TILE          Direct Submission
JOURNAL        Submitted (05-NOV-1998) Department Of Chemistry And Biochemistry,
              The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
              OK 73019, USA
REFERENCE       14 (bases 1 to 126312)
AUTHORS        Roe,B.A.
TILE          Direct Submission
JOURNAL        Submitted (21-NOV-1998) Department Of Chemistry And Biochemistry,
              The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
              OK 73019, USA
REFERENCE       15 (bases 1 to 126312)
AUTHORS        Roe,B.A.
TILE          Direct Submission
JOURNAL        Submitted (22-NOV-1998) Department Of Chemistry And Biochemistry,
              The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
              OK 73019, USA
REFERENCE       16 (bases 1 to 126312)
AUTHORS        Roe,B.A.
TILE          Direct Submission
JOURNAL        Submitted (13-SEP-1999) Department Of Chemistry And Biochemistry,
              The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
              OK 73019, USA
REFERENCE       17 (bases 1 to 126312)
AUTHORS        Roe,B.A.
TILE          Direct Submission
JOURNAL        Submitted (14-SEP-1999) Department Of Chemistry And Biochemistry,
              The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
              OK 73019, USA
REFERENCE       18 (bases 1 to 126312)
AUTHORS        Roe,B.A.
TILE          Direct Submission
JOURNAL        Submitted (15-SEP-1999) Department Of Chemistry And Biochemistry,
              The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
              OK 73019, USA
REFERENCE       19 (bases 1 to 126312)
AUTHORS        Roe,B.A.
TILE          Direct Submission
JOURNAL        Submitted (01-JUN-2000) Department Of Chemistry And Biochemistry,
              The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
              OK 73019, USA
COMMENT        On Sep 14, 1999 this sequence version replaced gi:5870977.
              Because these overlapping clones came from different libraries
              there are numerous instances of insertions, deletions, and single
              nucleotide polymorphisms in the overlapping regions below.
              AC000026(58b8) 8150 126312 (0) overlaps AC002059(p704f1059q13) 1
              118230 (54799) AC000026(58b8) 85607 126312 (0) overlaps
              L48038(e81f2) 1 40704 (0).
FEATURES       Location/Qualifiers
               1. .126312

```



```

sequence.
ACCESSION AL450164
VERSION AL450164.7 GI:12539770
KEYWORDS HTG
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 125051)
JOURNAL Dunn, M.
COMMENT Submitted (23-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonequery@sanger.ac.uk
On Jan 26, 2001 this sequence version replaced gi:12044654.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated repeat sequence elements. Where the sequence is
ambiguous, there is an annotation using the 'unsure' feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
This sequence
was generated from part of bacterial clone contigs of human
chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr6
RP11-550L8 is from the library RPI1-11.2 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6

IMPORTANT: This sequence is not the entire insert of clone
RP11-550L8 it may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.
The true left end of clone RPI-250C18 is at 124952 in this
sequence. The true right end of clone XX-AC002485 is at 100 in this
sequence.

FEATURES             Location/Qualifiers
     source           1..125051
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP11-550L8"
BASE COUNT          42749 a 21753 c 21024 g 39525 t
ORIGIN

Query Match          8.8%; Score 33.6; DB 9; Length 125051;
Best Local Similarity 55.0%; Pred. NO. 41;
Matches 66; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 37 tgaatctccgagccgagcgtgagtgctaccgcccgaagccgggaacaa 96
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 64537 TGCAGCACTCCAGAGTCCAAATGACTGAGCAGAGCTCCCTGCCAAAGACAGCAATAT 64478

QY 97 tatgaatcaggtattggcgccgcccacaataaagatttggataaagacgaatct 156
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 64477 TCATATTGGACTATTGGATCAGTCGACAGTATCTTTTATTATTAAACACCTGAATTT 64418

RESULT 12
AC055787
LOCUS
DEFINITION Homo sapiens chromosome 6 clone RP11-667G24 map 6, *** SEQUENCING

```

```

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

```

```

IN PROGRESS ***, 54 unordered pieces.
AC055787
AC055787.2 GI:8569182
HTG: HTGS_PHASE1.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 189278)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Homo sapiens chromosome 6, clone RP11-667G24
Unpublished
2 (bases 1 to 189278)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,
Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G.,
Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
Collamore, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S.,
Dodge, S., Domino, M., Doyle, M., Ferrel, P., FitzHugh, W., Gage, D.,
Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J.,
Levine, R., Liu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
Meldrum, J., Meneus, L., Mihova, P., Miranda, C., Mlenga, V., Morrow, J.,
Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Testa, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (18-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 16, 2000 this sequence version replaced gi:7582614.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WILBR
Web site: http://www.seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L6526
Center clone name: 667_G_24

```

```

TITLE
JOURNAL
COMMENT

```

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 54 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
1072 1171: gap of 100 bp
1172 2443: contig of 1272 bp in length
2444 2543: gap of 100 bp
2544 4110: contig of 1567 bp in length
4111 4210: gap of 100 bp
4211 5676: contig of 1466 bp in length
5677 5776: gap of 100 bp
5777 6954: contig of 1178 bp in length
6955 7054: gap of 100 bp
7055 8467: contig of 1413 bp in length
8468 8567: gap of 100 bp
8568 10046: contig of 1479 bp in length
10047 10146: gap of 100 bp
10147 11965: contig of 1819 bp in length

```


KEYWORDS

SOURCE
ORGANISM

influenza A virus (A/Swine/Quebec/192/81 (SwQc81))
influenza A virus (A/Swine/Quebec/192/81 (SwQc81))
Viruses; ssRNA negative-strand viruses; Orthomyxoviridae; Influenza
virus A and B group; Influenza A viruses; Influenza A virus;
unidentified subtype.

REFERENCE

AUTHORS
TITLE

Arora,D.J.S., Tijssen,P., Dea,S. and Henrichon,M.
Complete sequences of the neuraminidase genes of swine influenza
viruses (H1N1) associated with the respiratory disease in pigs
Virus Genes (1997) in press

JOURNAL

AUTHORS
TITLE

Dea,S., Arora,D.J.S., Tijssen,P. and Henrichon,M.
Direct Submission
Submitted (21-JAN-1997) CRV, Institut Armand-Frappier, 531 Boul.
des Prairies, Laval, QC H7N4Z3, Canada

FEATURES

source

1. 1438
/organism="influenza A virus (A/Swine/Quebec/192/81
(SwQc81))"
/strain="A/Sw/Quebec/192/81 (SwQc81)"
/db_xref="taxon:57280"
/note="tissue culture-adapted virus"

gene

1. 1438
/gene="NA"
1. 1410
/gene="NA"
/note="H1N1; peplomeric"
/codon_start=1
/product="neuraminidase"
/protein_id="AAD00583.1"
/db_xref="GI:4099319"

CDS

1. 1410
/translation="MTNQRIITIGTILIVGIVSLLLQIGNIVSLWISHTGTGKS
HPEVNSQVYVYNNVNTYVNTNIAAGQVTPPIIAGNSPLQIPISGMAIVSK
DNSIRIGKGDIVEMRPETSCSHFCRTFFLTGALLNDHRNSGTVKDRSPYILMS
CPIGEGAPNSYNSFESVANSASCHDGMGLIGTSGPDNGAVAVLKYGIITDIIS
WRNKLRTQSEVCINGSCFTITIDGSPNGQASYIKFMEKGIKISIELDAPNHY
EECSYPTDKVVCVCRDNHNSNRPWFSDQNLIDYIGYICSGFGDNPRSDNGKN
CGVLNSGANGVGFSEFRYNGVWIGRTKTSISRSFGFEMIDPWSGWTEDDSFSMKOD
IIALTDWGSYSGSFVQHPDLTGMNCRPCFWVELIRGQPKESTIWTSGSSISFCGWS
ETASWSWPGADLPFTIDK"

BASE COUNT
ORIGIN

454 a 258 c 340 g 386 t

Query Match
Best Local Similarity 8.7%; Score 33; DB 14; Length 1438;
Matches 54; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

Qy 104 cgagttatggcgggcccaataaagattccattggataagaacgaatctgttactt 163

Db 737 CAAGCAATGGCGAGCCCTCGTACAAAATATTCAAAATGGAGAAAGGAAGATTATTAAT 796

Qy 164 tgcttgagtcagcagcaaccatttccac 192

Db 797 CGATTGAAGTGGATGCACCTAAATTACCAC 825

RESULT 14
AC102175/c

LOCUS
AC102175
Mus musculus clone RP23-477G17, LOW-PASS SEQUENCE SAMPLING.

DEFINITION

AC102175

VERSION

AC102175.1 GI:17061261

KEYWORDS

HTG; HTGS, PHASE0.

SOURCE

house mouse.

ORGANISM

Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

AUTHORS
TITLE

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

JOURNAL

Unpublished

REFERENCE

AUTHORS

2 (bases 1 to 74854)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Bokhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepel,Y., Collangelo,M., Collins,S., Collamore,A., Cook,A.,
Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., LaRoque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

COMMENT

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L18168
Center clone name: 477_G_17

* NOTE: This record contains 90 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

* 1 712: contig of 712 bp in length
* 713 812: gap of 100 bp
* 813 1555: contig of 743 bp in length
* 1556 1655: gap of 100 bp
* 1656 2396: contig of 741 bp in length
* 2397 2496: gap of 100 bp
* 2497 3249: contig of 753 bp in length
* 3250 3349: gap of 100 bp
* 3350 4063: contig of 714 bp in length
* 4064 4163: gap of 100 bp
* 4164 4893: contig of 730 bp in length
* 4894 4993: gap of 100 bp
* 4994 5689: contig of 696 bp in length
* 5690 5789: gap of 100 bp
* 5790 6539: contig of 750 bp in length
* 6540 6639: gap of 100 bp
* 6640 7386: contig of 747 bp in length
* 7387 7486: gap of 100 bp
* 7487 8199: contig of 713 bp in length
* 8200 8299: gap of 100 bp
* 8300 9019: contig of 720 bp in length
* 9020 9119: gap of 100 bp
* 9120 9852: contig of 733 bp in length
* 9853 9952: gap of 100 bp

* 9953 10603: contig of 651 bp in length
* 10604 10703: gap of 100 bp
* 10704 11455: contig of 752 bp in length
* 11456 11555: gap of 100 bp
* 11556 12282: contig of 727 bp in length
* 12283 12382: gap of 100 bp
* 12383 13126: contig of 744 bp in length
* 13127 13226: gap of 100 bp
* 13227 13922: contig of 696 bp in length
* 13923 14022: gap of 100 bp
* 14023 14755: contig of 733 bp in length
* 14756 14855: gap of 100 bp
* 14856 15599: contig of 744 bp in length
* 15600 15699: gap of 100 bp
* 15700 16435: contig of 736 bp in length
* 16436 16535: gap of 100 bp
* 16536 17265: contig of 730 bp in length
* 17266 17365: gap of 100 bp
* 17366 18082: contig of 717 bp in length
* 18083 18182: gap of 100 bp
* 18183 18909: contig of 727 bp in length
* 18910 19009: gap of 100 bp
* 19010 19713: contig of 710 bp in length
* 19720 19819: gap of 100 bp
* 19820 20556: contig of 737 bp in length
* 20557 20656: gap of 100 bp
* 20657 21396: contig of 740 bp in length
* 21397 21496: gap of 100 bp
* 21497 22238: contig of 742 bp in length
* 22239 22338: gap of 100 bp
* 22339 23068: contig of 730 bp in length
* 23069 23168: gap of 100 bp
* 23169 23894: contig of 726 bp in length
* 23895 23994: gap of 100 bp
* 23995 24733: contig of 739 bp in length
* 24734 24833: gap of 100 bp
* 24834 25567: contig of 734 bp in length
* 25568 25667: gap of 100 bp
* 25668 26394: contig of 727 bp in length
* 26395 26494: gap of 100 bp
* 26495 27239: contig of 745 bp in length
* 27240 27339: gap of 100 bp
* 27340 28073: contig of 734 bp in length
* 28074 28173: gap of 100 bp
* 28174 28923: contig of 750 bp in length
* 28924 29023: gap of 100 bp
* 29024 29753: contig of 730 bp in length
* 29754 29853: gap of 100 bp
* 29854 30594: contig of 741 bp in length
* 30595 30694: gap of 100 bp
* 30695 31431: contig of 737 bp in length
* 31432 31531: gap of 100 bp
* 31532 32252: contig of 721 bp in length
* 32253 32352: gap of 100 bp
* 32353 33097: contig of 745 bp in length
* 33098 33197: gap of 100 bp
* 33198 33949: contig of 752 bp in length
* 33950 34049: gap of 100 bp
* 34050 34781: contig of 732 bp in length
* 34782 34881: gap of 100 bp
* 34882 35598: contig of 717 bp in length
* 35599 35698: gap of 100 bp
* 35699 36438: contig of 740 bp in length
* 36439 36538: gap of 100 bp
* 36539 37289: contig of 751 bp in length
* 37290 37389: gap of 100 bp
* 37390 38086: contig of 697 bp in length
* 38087 38186: gap of 100 bp
* 38187 38916: contig of 730 bp in length
* 38917 39016: gap of 100 bp
* 39017 39763: contig of 747 bp in length
* 39764 39863: gap of 100 bp
* 39864 40616: contig of 753 bp in length

* 40617 40716: gap of 100 bp
* 40717 41450: contig of 734 bp in length
* 41451 41550: gap of 100 bp
* 41551 42274: contig of 724 bp in length
* 42275 42374: gap of 100 bp
* 42375 43106: contig of 732 bp in length
* 43107 43206: gap of 100 bp
* 43207 43960: contig of 754 bp in length
* 43961 44060: gap of 100 bp
* 44061 44814: contig of 754 bp in length
* 44815 44914: gap of 100 bp
* 44915 45641: contig of 727 bp in length
* 45642 45741: gap of 100 bp
* 45742 46462: contig of 721 bp in length
* 46463 46562: gap of 100 bp
* 46563 47306: contig of 744 bp in length
* 47307 47406: gap of 100 bp
* 47407 48135: contig of 729 bp in length
* 48136 48235: gap of 100 bp
* 48236 48972: contig of 737 bp in length
* 48973 49072: gap of 100 bp
* 49073 49807: contig of 735 bp in length
* 49808 49907: gap of 100 bp
* 49908 50639: contig of 732 bp in length
* 50640 50739: gap of 100 bp
* 50740 51474: contig of 735 bp in length
* 51475 51574: gap of 100 bp
* 51575 52306: contig of 732 bp in length
* 52307 52406: gap of 100 bp
* 52407 53137: contig of 731 bp in length
* 53138 53237: gap of 100 bp
* 53238 53953: contig of 716 bp in length
* 53954 54053: gap of 100 bp
* 54054 54775: contig of 722 bp in length
* 54776 54875: gap of 100 bp
* 54876 55593: contig of 718 bp in length
* 55594 55693: gap of 100 bp
* 55694 56418: contig of 725 bp in length
* 56419 56518: gap of 100 bp
* 56519 57267: contig of 749 bp in length
* 57268 57367: gap of 100 bp

Query Match 8.7%; Score 33; DB 2; Length 74854;

Best Local Similarity 49.7%; Pred No. 61;

Matches 84; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 29 acatcttatgaatctctgccaggccgaagcctggaggtgtaccgcgcgaagcc 88

Db 57085 AATACCTATTATTTCTCCAGATAGCCAAAGGCTGTGAAGGCCACGAGTTCCACAGAGA 57026

QY 89 gggacaataatgaatcagattatggcgcccaataaaagattccattggataaga 148

Db 57025 GAGAAAAGGATGCATGCTATTTGCCAGGTGCTCAGCAAGTTTCACGCACAAATCTGGGAA 56966

QY 149 acgaatctgtacttgcgttcgagtcgacgcaaccacttccacgtcac 197

Db 56965 TCACCATAGCTGCATGATGCTTTCCACGGATCTAATTACCACTTCAC 56917

RESULT 15

AC025672

LOCUS Homo sapiens chromosome 18 clone RP11-715F3 map 18, WORKING DRAFT
DEFINITION SEQUENCE, 20 unordered pieces.

ACCESSION AC025672

VERSION AC025672.2 GI:10198540

KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.

SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 167178)

AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.

TITLE
JOURNAL
REFERENCE
AUTHORS

Homo sapiens chromosome 18, clone RP11-715F3
Unpublished
2 (bases 1 to 167178)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,
Boguslavsky, L., Bouckgalter, B., Brown, A., Burkett, G.,
Campiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
Collamore, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J.S.,
Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D.,
Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., LaRoque, K., Lamazares, R., Landers, T., Lehoczy, J.,
Levine, R., Liu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
McLirim, J., Meneus, L., Mihova, T., Miranda, C., Mieng, V., Morrow, J.,
Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neil, D., Olivari, T.M., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tessfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (12-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 20, 2000 this sequence version replaced gi:7230282.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: 715_F-3
Center clone name: 715_F-3
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 159523 bases at least Q40
Consensus quality: 163250 bases at least Q30
Consensus quality: 164626 bases at least Q20
Insert size: 176000; agarose-fp
Quality coverage: 4.4 in Q20 bases; agarose-fp
Quality coverage: 4.7 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 20 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 27917: contig of 27917 bp in length
* 27918 28017: gap of 100 bp
* 28018 29021: contig of 1004 bp in length
* 29022 29121: gap of 100 bp
* 29122 30937: contig of 1816 bp in length
* 30938 31037: gap of 100 bp
* 31038 33054: contig of 2017 bp in length
* 33055 33154: gap of 100 bp
* 33155 34403: contig of 1249 bp in length
* 34404 34503: gap of 100 bp
* 34504 36545: contig of 2042 bp in length
* 36546 36645: gap of 100 bp
* 36646 39583: contig of 2938 bp in length
* 39584 39683: gap of 100 bp

TITLE
JOURNAL
COMMENT

* 39684 44557: contig of 4874 bp in length
* 44558 44657: gap of 100 bp
* 44658 48993: contig of 4336 bp in length
* 48994 49093: gap of 100 bp
* 49094 53867: contig of 4774 bp in length
* 53868 53967: gap of 100 bp
* 53968 58665: contig of 4698 bp in length
* 58666 58765: gap of 100 bp
* 58766 65551: contig of 6786 bp in length
* 65552 65651: gap of 100 bp
* 65652 72916: contig of 7265 bp in length
* 72917 73016: gap of 100 bp
* 73017 83269: contig of 10253 bp in length
* 83270 83369: gap of 100 bp
* 83370 93557: contig of 10188 bp in length
* 93558 93657: gap of 100 bp
* 93658 103579: contig of 9922 bp in length
* 103580 103679: gap of 100 bp
* 103680 117213: contig of 13534 bp in length
* 117214 117313: gap of 100 bp
* 117314 133144: contig of 15831 bp in length
* 133145 133244: gap of 100 bp
* 133245 160275: contig of 27031 bp in length
* 160276 160375: gap of 100 bp
* 160376 167178: contig of 6803 bp in length.
FEATURES
Location/Qualifiers
1..167178
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18"
/clone="RP11-715F3"
/clone_lib="RPC1-11 Human Male BAC"
1..27917
/note="assembly_fragment"
clone_end:SP6
vector_side:left
28018..29021
/note="assembly_fragment"
29122..30937
/note="assembly_fragment"
31038..33054
/note="assembly_fragment"
33155..34403
/note="assembly_fragment"
34504..36545
/note="assembly_fragment"
36646..39583
/note="assembly_fragment"
39684..44557
/note="assembly_fragment"
44558..48993
/note="assembly_fragment"
49094..53867
/note="assembly_fragment"
53968..58665
/note="assembly_fragment"
58766..65551
/note="assembly_fragment"
65652..72916
/note="assembly_fragment"
73017..83269
/note="assembly_fragment"
83370..93557
/note="assembly_fragment"
93658..103579
/note="assembly_fragment"
103680..117213
/note="assembly_fragment"
117314..133144
/note="assembly_fragment"
133245..160275
/note="assembly_fragment"

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature 160376..167178
/note="assembly_fragment
clone_end:T7
vector_slide:right"

BASE COUNT 52172 a 31878 c 31581 g 49646 t 1901 others
ORIGIN

Query Match 8.7%; Score 33; DB 2; Length 167178;
Best Local Similarity 54.5%; Pred. No. 61;
Matches 66; Conservative 0; Mismatches 55; Indels 0; Gaps 0;
QY 99 tgaatcgagttatggcgggccacacataaaagattccatttggataaagaacgaatctgt 158
DB 15726 TGAAGAAGAGCTAGGGCCTTCTCTGAATTAGACTTTGGATTAGGGAATCTTATGGCTGG 15785
QY 159 tactttgcttgacgtgaacgaaccactttccacgtccccaatccaggtagtagcttgc 218
DB 15786 TTTGATCTATCGAGACCACACAAACGTTCTCCATGTCTCAGCAATAAAGCTGTTTGTTC 15845
QY 219 t 219
DB 15846 T 15846

Search completed: June 19, 2002, 15:48:13
Job time: 13751 sec


```

XX Kimmerly WJ;
XX WPI; 2001-316495/33.
XX Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
XX useful for vaccinating against infections, e.g. endocarditis -
XX Claim 8; Page 1285-1287; 2188pp; English.
XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
XX (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
XX (I) and (II) can have antibacterial activity and therefore can be used
XX in vaccination. The nucleic acids (I) may be used to produce the
XX S. epidermidis polypeptides (II) via the production of vectors
XX containing them which are used to produce hosts cells which express the
XX polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
XX used to vaccinate subjects and to raise antibodies against the bacteria.
XX The polypeptides may also be used to assay for other inhibitors of their
XX activity and therefore identify compounds that may be used for the
XX treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
XX AAH5090 represent specifically claimed S. epidermidis genomic DNA
XX polynucleotide sequences from the present invention. AAH55091 to
XX AAH55098 represent oligonucleotide sequences and primers which are used
XX in the exemplification of the present invention.
XX N.B. The present invention specifically claims all the polynucleotide
XX sequences given in the sequence listing of the present specification,
XX however the sequence listing only goes up to SEQ ID NO:4454 so even
XX though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
XX no sequences are present for SEQ ID NO:4455 to 4464.
XX Sequence 3295 BP; 971 A; 667 C; 456 G; 1201 T; 0 other;

Query Match      8.5%; Score 32.4; DB 22; Length 3295;
Best Local Similarity 53.1%; Pred. No. 0.88;
Matches 69; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 200 atccaggtgagtagcttgctggagaaagcgcgaagcctctatccgcgtccattgt 259
   || |||| || || || || || || || || || || || || || || || || ||
DB 776 attaaagtattattcttagtatagaaaatgcctgagacaatgattgcccgacattta 835

QY 260 ttitgcgcatcgacggcgtgattgatctgcgccaaaacctctctgaagtcogtgc 319
   || || || || || || || || || || || || || || || || || || ||
DB 836 ttattggcaattgacacccgagtggaactgggctaattctcaattttcaactatgccc 895

QY 320 taagtataaa 329
   || || || ||
DB 896 tcattaccaa 905

RESULT 4
AAH54750
ID AAH54750 standard; DNA; 3518 BP.
XX
XX AAH54750;
XX
XX 03-SEP-2001 (first entry)
XX
XX S. epidermidis genomic polynucleotide sequence SEQ ID NO:4114.
XX
XX Staphylococcus epidermidis SRI strain; infection; diagnosis;
XX KW vaccination; endocarditis; ds.
XX
XX Staphylococcus epidermidis.
XX OS
XX WO200134809-A2.
XX PN
XX 17-MAY-2001.
XX PD
XX 09-NOV-2000; 2000WO-US30782.
XX PF
XX 09-NOV-1999; 99US-0164258.
XX PR

(GLAX ) GLAXO GROUP LTD.
XX Kimmerly WJ;
XX WPI; 2001-316495/33.
XX Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
XX useful for vaccinating against infections, e.g. endocarditis -
XX Claim 8; Page 1808-1809; 2188pp; English.
XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
XX (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
XX (I) and (II) can have antibacterial activity and therefore can be used
XX in vaccination. The nucleic acids (I) may be used to produce the
XX S. epidermidis polypeptides (II) via the production of vectors
XX containing them which are used to produce hosts cells which express the
XX polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
XX used to vaccinate subjects and to raise antibodies against the bacteria.
XX The polypeptides may also be used to assay for other inhibitors of their
XX activity and therefore identify compounds that may be used for the
XX treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
XX AAH5090 represent specifically claimed S. epidermidis genomic DNA
XX polynucleotide sequences from the present invention. AAH55091 to
XX AAH55098 represent oligonucleotide sequences and primers which are used
XX in the exemplification of the present invention.
XX N.B. The present invention specifically claims all the polynucleotide
XX sequences given in the sequence listing of the present specification,
XX however the sequence listing only goes up to SEQ ID NO:4454 so even
XX though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
XX no sequences are present for SEQ ID NO:4455 to 4464.
XX Sequence 3518 BP; 1091 A; 677 C; 495 G; 1255 T; 0 other;

Query Match      8.5%; Score 32.4; DB 22; Length 3518;
Best Local Similarity 53.1%; Pred. No. 0.91;
Matches 69; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 200 atccaggtgagtagcttgctggagaaagcgcgaagcctctatccgcgtccattgt 259
   || |||| || || || || || || || || || || || || || || || || ||
DB 2590 attaaagtattattcttagtatagaaaatgcctgagacaatgattgcccgacattta 2649

QY 260 ttitgcgcatcgacggcgtgattgatctgcgccaaaacctctctgaagtcogtgc 319
   || || || || || || || || || || || || || || || || || || ||
DB 2650 ttattggcaattgacacccgagtggaactgggctaattctcaattttcaactctagccc 2709

QY 320 taagtataaa 329
   || || || ||
DB 2710 tcattaccaa 2719

RESULT 5
AAH54492
ID AAH54492 standard; DNA; 3985 BP.
XX
XX AAH54492;
XX
XX 03-SEP-2001 (first entry)
XX
XX S. epidermidis genomic polynucleotide sequence SEQ ID NO:3856.
XX DE
XX Staphylococcus epidermidis SRI strain; infection; diagnosis;
XX KW vaccination; endocarditis; ds.
XX
XX Staphylococcus epidermidis.
XX OS
XX WO200134809-A2.
XX PN
XX 17-MAY-2001.
XX PD
XX 09-NOV-2000; 2000WO-US30782.
XX PF
XX 09-NOV-2000; 2000WO-US30782.
XX PR

```


AAH99166 to AAH99904 encode the human proteins given in AAH25225 to AAH25963. The proteins can have activities based on the tissues and cells they are expressed in, such as: antinflammatory; antirheumatic; antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant; central nervous system; viricide; anti-HIV; fungicide; antimutagen; cardiovascular; antihaemic; antiagregant; haemostatic; vulnery; anticulcer; osteopathic; dermatological; antiageric; antiasthmatic; antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic; antiparkinsonian; and immunostimulant. The proteins and polynucleotides encoding them can be used in gene therapy, antisense therapy and vaccing production. The proteins and polynucleotides are useful for screening for agonists or antagonists of a protein and for the treatment and diagnosis

SQ Sequence 5653 BP; 1262 A; 1680 C; 1632 G; 1079 T; 0 other;

PS CLAIM 1; SEQ ID NO 4843; 100/6pp; English.

XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with neurotropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.

Sequence 5653 BP; 1262 A; 1680 C; 1632 G; 1079 T; 0 other;

Query Match 8.5%; Score 32.4; DB 22; Length 5653;
 Best Local Similarity 55.3%; Pred. No. 1.1;
 Matches 63; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 4 cccgcccgcacgcgtatcgtttacatcttataatcctcctccagggccgagcct 63
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 DB 770 CCCAGAGCCTCCCATGTCTCTGCTGCTGAGAAATGCCCTCCCATGCCGTGAGGT 711
 QY 64 gggaggtgctaccgcccagggcgggaaacaatatgaatcgagttatggcgg 117
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 DB 710 GTTAGGTGTTAGGGCCCAAGGGGAAACCACTTGAGTCTTGTTGTGTGG 657

RESULT 9

AAH99496
 ID AAH99496 standard; cDNA; 5654 BP.

XX AAH99496;

DT 16-OCT-2001 (first entry)

DE Human protein encoding cDNA sequence SEQ ID NO:331.

XX Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
 KW antiinflammatory; anti-rheumatic; antiarthritic; immunosuppressive;
 KW antibacterial; endocrine; cardiant; cardiant; central nervous system; virucide;
 KW anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;
 KW antiaggregant; haemostatic; vulnerary; antiulcer; osteopathic; eczema;
 KW dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;
 KW neuroprotective; antidepressant; neurotropic; antiparkinsonian; infection;
 KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
 KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
 KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autolmmunity;
 KW genetic disease; haematopoietic disorder; platelet disorder; asthma;
 KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;
 KW allergic rhinitis; diabetes; multiple sclerosis; depression;
 KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
 KW neurological disorder; ss.

OS Homo sapiens.

XX WO200153455-A2.

PN 26-JUL-2001.

XX 22-DEC-2000; 2000WO-US35017.

XX 23-DEC-1999; 99US-0471275.

PR 21-JAN-2000; 2000US-0488725.

PR 25-APR-2000; 2000US-0552317.

XX (HYSE-) HYSEQ INC.

XX

PI Tang YT, Liu C, Drmanac RT;

DR WPI: 2001-457603/49.

DR P-PSDB; AAM25555.

XX Isolated human polynucleotides encoding polypeptides, useful for the
 CC treatment and diagnosis of e.g. cancer, ulcers and HIV infection -
 PT Claim 1; Page 444-446; 1217pp; English.

XX AAH99166 to AAH99904 encode the human proteins given in AAM25225 to
 CC AAM25963. The proteins can have activities based on the tissues and
 CC cells they are expressed in, such as: antiinflammatory; antirheumatic;
 CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant;
 CC central nervous system; virucide; anti-HIV; fungicide; antimutagen;
 CC cardiovascular; antianaemic; antiaggregant; haemostatic; vulnerary;
 CC antiulcer; osteopathic; dermatological; antiallergic; antiasthmatic;
 CC antidiabetic; cytostatic; neuroprotective; antidepressant; neurotropic;
 CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
 CC encoding them can be used in gene therapy, antisense therapy and vaccine
 CC production. The proteins and polynucleotides are useful for screening for
 CC agonists or antagonists of a protein and for the treatment and diagnosis
 CC of disorders associated with the activity of a protein e.g. inflammation,
 CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
 CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
 CC infections, autoimmunity, genetic diseases, haematopoietic disorders,
 CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
 CC osteoporosis, severe combined immunodeficiency, eczema, allergic
 CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
 CC Alzheimer's disease, Parkinson's disease, neurodegenerative and
 CC neurological disorders.

XX Sequence 5654 BP; 1078 A; 1631 C; 1682 G; 1263 T; 0 other;

Query Match 8.5%; Score 32.4; DB 22; Length 5654;

Best Local Similarity 55.3%; Pred. No. 1.1;

Matches 63; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 4 cccgcccgcacgcgtatcgtttacatcttataatcctcctccagggccgagcct 63
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 DB 4887 ccagagcctcccatgctcctgctgtagaattgcccctccatgcgcgtgaggt 4946

QY 64 gggaggtgctaccgcccagggcgggaaacaatatgaatcgagttatggcgg 117
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

DB 4947 gttaggtggttagggcccaagggaaccacttgactgtgtgtgtgg 5000

RESULT 10

ABAO8796

ID ABAO8796 standard; cDNA; 5665 BP.

XX ABAO8796;

DT 11-JAN-2002 (first entry)

DE Human d889M15.3 homologue-encoding cDNA, SEQ ID NO:572.

XX Human; cytokine; cell proliferation; cell differentiation; growth factor;
 KW haematopoiesis regulation; tissue growth; immunomodulator; activin;
 KW inhibin; chemotaxis; chemokinesis; thrombolytic; oncogenesis;
 KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;
 KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
 KW chronic inflammatory condition; proliferative retinopathy;
 KW atherosclerosis; coronary heart disease; arterial ischaemia;
 KW bone disorder; osteoporosis; vascular growth disorder;
 KW tissue regeneration; wound healing; infection; immune disorder;
 KW cell culture; drug screening; gene therapy; antiinflammatory;
 KW antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;
 KW cytostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;
 KW antifungal; vulnerary; antiulcer; ss.


```
Query Match      8.2%; Score 31.2; DB 22; Length 2497;
Best Local Similarity 50.0%; Pred. No. 2.1;
Matches 78; Conservative 0; Mismatches 78; Indels 0; Gaps 0;
```


Result No.	Category	Query			DB	ID	Description
		Match	Length	Score			
1	C	291	76.4	1291	4	US-09-462-975-1	Sequence 1, Appli
2		32.2	8.5	4411529	4	US-09-103-840A-1	Sequence 1, Appli
3		31	8.1	31	4	US-09-462-975-2	Sequence 2, Appli
4	C	30	7.9	265	3	US-08-289-222E-5	Sequence 5, Appli
5		30	7.9	265	4	US-09-054-526B-5	Sequence 3, Appli
6		30	7.9	2088	4	US-09-351-414-3	Sequence 5, Appli
7	C	30	7.9	2272	1	US-08-482-577B-1	Sequence 1, Appli
8		30	7.9	2272	3	US-08-289-222E-2	Sequence 2, Appli
9		30	7.9	2272	4	US-09-218-176-1	Sequence 1, Appli
10	C	30	7.9	2272	4	US-09-054-526B-2	Sequence 2, Appli
11		29.4	7.7	1414	4	US-08-686-968C-228	Sequence 228, App
12		27.6	7.3	3129	4	US-09-387-695-1	Sequence 1, Appli
13	C	27.6	7.2	1128	4	US-08-976-259-96	Sequence 96, Appli
14		27.2	7.08	1	US-08-078-090-24	Sequence 24, Appli	
15		27.4	7.2	1065	1	US-08-249-554-1	Sequence 24, Appli
16	C	27.4	7.2	1065	1	US-08-249-555A-1	Sequence 1, Appli
17		27.4	7.2	1065	1	US-08-249-584-1	Sequence 1, Appli
18		27.4	7.2	1065	1	US-08-734-752-1	Sequence 1, Appli
19	C	27.4	7.2	1065	1	US-08-078-090-1	Sequence 1, Appli
20		27.4	7.2	1065	5	PTC-US95-03628-1	Sequence 1, Appli
21		27.4	7.2	1065	5	PTC-US95-03789-1	Sequence 1, Appli
22	C	27.4	7.2	2052	1	US-08-655-836-2	Sequence 2, Appli
23		27.4	7.2	2202	2	US-09-020-753-2	Sequence 2, Appli
24		27.4	7.2	2364	1	US-08-655-836-3	Sequence 3, Appli
25	C	27.4	7.2	2364	2	US-09-020-753-3	Sequence 3, Appli
26		27.4	7.2	2799	3	US-08-968-752B-5	Sequence 5, Appli
27		27.4	7.2	3027	4	US-08-818-112-110	Sequence 110, App

Qy 241 ctatctaccgtccattgttttttgcgcgacgacgagctgagttgattctgg 291
|||||
Db 1241 ctatctaccgtccattgttttttgcgcgacgacgagctgagttgattctgg 1291
|||||

RESULT 2

US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match 8.58; Score 32.2; DB 4; Length 4411529;
Best Local Similarity 57.4%; Pred. No. 4.4;
Matches 58; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

Qy 208 gagttagctgtcgtgagaagaacgcgaagcctctatctaccgtccattgttttgcgc 267
|||||
Db 3577728 GAGCGGCACGGCGGAACAACACCGGAGCTGCCGTCTGCGGATCGGTGTTCAC 3577669
|||||

Qy 268 gatcgagcggctgagttgatcgtgcgcacaaacctctgt 308
|||||

Db 3577668 GCAGTGTTCGCGGCGGAGGAGGATCCCACTGCT 3577628
|||||

RESULT 3

US-09-462-975-2/c
; Sequence 2, Application US/09462975
; Patent No. 6303345
; GENERAL INFORMATION:
; APPLICANT: Rohde, Wolfgang
; APPLICANT: Becker, Dieter
; APPLICANT: Randles, John W.
; APPLICANT: Hehn, Alain
; APPLICANT: Salamini, Francesco
; TITLE OF INVENTION: USE OF VIRUS DNA AS A PROMOTER
; FILE REFERENCE: 23232.000301
; CURRENT APPLICATION NUMBER: US/09/462,975
; CURRENT FILING DATE: 2000-05-17
; PRIOR APPLICATION NUMBER: PCT/EP98/04345
; PRIOR FILING DATE: 1998-07-13
; PRIOR APPLICATION NUMBER: 19730502.4
; PRIOR FILING DATE: 1997-07-16
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 31
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: /note-synthetic
US-09-462-975-2

Query Match 8.18; Score 31; DB 4; Length 31;

Best Local Similarity 100.0%; Pred. No. 0.019;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 331 agcccgagggttagtattaccgccgcgcgcct 361
|||||
Db 31 AGCCCGGGGGCTAGTATTACCCCGCGGCT 1
|||||

RESULT 4

US-08-289-222E-5
; Sequence 5, Application US/08289222E
; Patent No. 6120760
; GENERAL INFORMATION:
; APPLICANT: HOTTEN, GERTRUD
; APPLICANT: NEIDHARDT, HELGE
; APPLICANT: BECHTOLD, ROLF
; APPLICANT: POHL, JENS
; TITLE OF INVENTION: GROWTH/DIFFERENTIATION FACTORS OF THE TGF-B
; TITLE OF INVENTION: FAMILY
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIKAIKO, MARCELSTEIN, MURRAY & ORAM
; STREET: 655 FIFTEENTH STREET, N. W., G STREET LOBBY,
; CITY: SUITE 330
; STATE: WASHINGTON
; COUNTRY: USA
; ZIP: 20005-5701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/289,222E
; FILING DATE: 25-AUG-1999
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/289,222
; FILING DATE: 12-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE P 44 23 190.3
; FILING DATE: 07-JUL-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EPO 92102324.8
; FILING DATE: 12-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP93/00350
; FILING DATE: 12-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: KITTS, MONICA CHIN
; REGISTRATION NUMBER: 36,105
; REFERENCE/DOCKET NUMBER: P564-9021
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202/638-5000
; TELEFAX: 202/638-4810
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 265 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA from mRNA
US-08-289-222E-5

Query Match 7.9%; Score 30; DB 3; Length 265;

Best Local Similarity 59.3%; Pred. No. 0.14;
Matches 51; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

Qy 292 cgccaaaacctctgcttaagtccgtccgttaagataaataagcgggggtagtattac 351
|||||
Db 19 CGCGCATGAAGTCTCTGCATAGGGCAGTCCCACTACACATAGCAGGCATGCCTGGTATTGC 78
|||||

Qy 352 cccgcggctcccccaacctctgctaa 377
||| | | | | | | | |
Db 79 TGCCTCCCTTTCACACTGCAGTGTCTCA 104

```

RESULT      5
US-09-054-526B-5
: Sequence 5, Application US/09054526B
: Patent No. 6197550
: GENERAL INFORMATION:
: APPLICANT: H TTEN, GERTRUD
: APPLICANT: NEIDHARDT, HELGE
: APPLICANT: BECHTOLD, ROLF
: APPLICANT: POHL, JENS
: TITLE OF INVENTION: DNA SEQUENCES ENCODING NOVEL
: TITLE OF INVENTION: GROWTH/DIFFERENTIATION FACTORS
: NUMBER OF SEQUENCES: 53
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: NIKAIIDO, MARMELESTEIN, MURRAY & ORAM LLP
: STREET: 655 FIFTEENTH STREET, N. W., G STREET LOBBY,
: CITY: WASHINGTON
: STATE: DC
: COUNTRY: USA
: ZIP: 20005-5701
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC Compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/054,526B
: FILING DATE: 03-APR-1998
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/289,222
: FILING DATE: 12-AUG-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: DE P 44 23 190.3
: FILING DATE: 01-JUL-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: EPO 92102324.8
: FILING DATE: 12-FEB-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/EP93/00350
: FILING DATE: 12-FEB-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: KITTS, MONICA CHIN
: REGISTRATION NUMBER: 36,105
: REFERENCE/DOCKET NUMBER: P564-8005
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202/638-5000
: TELEFAX: 202/638-4810
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 265 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: both
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA from mRNA
US-09-054-526B-5

```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EPO 92102324.8
; FILING DATE: 12-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP93/00350
; FILING DATE: 12-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: KITS, MONICA CHIN
; REGISTRATION NUMBER: 36,105
; REFERENCE/DOCKET NUMBER: P564-9021
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202/638-5000
; TELEFAX: 202/638-4810
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2272 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA from mRNA
;
US-08-289-222E-2

Query Match 7.9%; Score 30; DB 3; Length 2272;
Best Local Similarity 59.3%; Pred. No. 0.47;
Matches 51; Conservative 0; Mismatches 35; Indels 0; Gaps

QY 292 cgcacaaacctgtctaagtcgccgtgctgaataataagccgcgggctagtagttacc 351
||||| ||| ||||| || ||||| || ||||| || ||||| || ||||| ||
Db 940 CGCATGAACCTTCGATAGGGCAGTGCACCATACATAGCAGGCATGCCTGCTATTGC 999
QY 352 cccgcggctcccaacctctgctaa 377
Db 1000 TGCTCTTTACACTGCAGTGCTCA 1025

RESULT 9
US-09-218-176-1
; Sequence 1, Application US/09218176
; Patent No. 6171584
; GENERAL INFORMATION:
; APPLICANT: H TIEN, Gertrud
; APPLICANT: NEIDHARDT, Helge
; APPLICANT: BECHTOLD, Rolf
; APPLICANT: POHL, Jens
; APPLICANT: PAULISTA, Michael
; TITLE OF INVENTION: NEW GROWTH/DIFFERENTIATION FACTORS OF THE
; TITLE OF INVENTION: TGF- FAMILY
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIKAIIDO, MARCELSTEIN, MURRAY & ORAM LLP
; STREET: 655 Fifteenth Street, N. W., G Street Lobby,
; STREET: Suite 330
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/218,176
; FILING DATE: Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/679,048
; FILING DATE: 12-JUL-1996
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP96/03065
; FILING DATE: 12-JUL-1996

```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP93/00350
; FILING DATE: 2-FEB-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/482,577
; FILING DATE: 7-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 92 102 324.8
; FILING DATE: 12-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE P 44 23 190.3
; FILING DATE: 01-JUL-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP93/00350
; FILING DATE: 12-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: KITTS, Monica Chin
; REGISTRATION NUMBER: 36,105
; REFERENCE/DOCKET NUMBER: P564-6010
; TELEPHONE: 202/638-5000
; TELEFAX: 202/638-4810
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2272 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
US-09-218-176-1

```

```

Query Match          7.9%; Score 30; DB 4; Length 2272;
Best Local Similarity 59.3%; Pred. No. 0.47;
Matches 51; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 292 cgccaaaaacctctgtaagtcctgctgaatataaataagcgggggctagtattac 351
      ||||| ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 940 CGCCATGAACCTTCTGCATAGGCGAGTGCCCACTACACATAGCAGCATGCGCTGTATTGC 999

QY 352 ccccggggctcccaacctctgctaa 377
      || || || || || || || || || || || || || || || || || || || || || ||
Db 1000 TGCCTCCTTTTCACACTGCAGTGCTCA 1025

```

```

RESULT 10
US-09-054-526B-2
; Sequence 2, Application US/09054526B
; Patent No. 6197550
; GENERAL INFORMATION:
; APPLICANT: H TTEN, GERTRUD
; APPLICANT: NEIDHARDT, HELGE
; APPLICANT: BECHTOLD, ROLF
; APPLICANT: POHL, JENS
; TITLE OF INVENTION: DNA SEQUENCES ENCODING NOVEL
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIKAIKO, MARCELSTEIN, MURRAY & ORAM LLP
; STREET: 655 FIFTEENTH STREET, N. W., G STREET LOBBY,
; STREET: SUITE 330
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-5701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/054,526B
; FILING DATE: 03-APR-1998

```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/289,222
; FILING DATE: 12-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE P 44 23 190.3
; FILING DATE: 01-JUL-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EPO 92102324.8
; FILING DATE: 12-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP93/00350
; FILING DATE: 12-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: KITTS, Monica Chin
; REGISTRATION NUMBER: 36,105
; REFERENCE/DOCKET NUMBER: P564-8005
; TELEPHONE: 202/638-5000
; TELEFAX: 202/638-4810
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2272 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA from mRNA
US-09-054-526B-2

```

```

Query Match          7.9%; Score 30; DB 4; Length 2272;
Best Local Similarity 59.3%; Pred. No. 0.47;
Matches 51; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 292 cgccaaaaacctctgtaagtcctgctgaatataaataagcgggggctagtattac 351
      ||||| ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 940 CGCCATGAACCTTCTGCATAGGCGAGTGCCCACTACACATAGCAGCATGCGCTGTATTGC 999

QY 352 ccccggggctcccaacctctgctaa 377
      || || || || || || || || || || || || || || || || || || || || || ||
Db 1000 TGCCTCCTTTTCACACTGCAGTGCTCA 1025

```

```

RESULT 11
US-08-686-968C-228
; Sequence 228, Application US/08686968C
; Patent No. 6221361
; GENERAL INFORMATION:
; APPLICANT: Cochran, Mark D.
; APPLICANT: Junker, David E.
; TITLE OF INVENTION: Recombinant Swinepox Virus
; FILE REFERENCE: 39119-H/JML
; CURRENT APPLICATION NUMBER: US/08/686,968C
; CURRENT FILING DATE: 1996-07-25
; NUMBER OF SEQ ID NOS: 231
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 228
; LENGTH: 1414
; TYPE: DNA
; ORGANISM: Swinepox virus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1407)
US-08-686-968C-228

```

```

Query Match          7.7%; Score 29.4; DB 4; Length 1414;
Best Local Similarity 56.8%; Pred. No. 0.61;
Matches 54; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 98 atgaatcgagttatggcgggggcccaataaagaattccatttggataagaacgaatctg 157
      || | || || || || || || || || || || || || || || || || || || || ||
Db 731 atggcccaagcaatgggcaagcctctgcatacaaatattcaaatggagaaagggaagatta 790

```


REFERENCE/DOCKET NUMBER: BERGSTROM2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)628-5197
TELEFAX: (202)737-3528
TELEX: 248633
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 708 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-078-090-24

Query Match 7.2%; Score 27.4; DB 1; Length 708;
Best Local Similarity 55.9%; Pred. No. 2.3;
Matches 52; Conservative 0; Mismatches 41; Indels 0; Gaps 0;
QY 12 cccacgcgtatcggtttacatcttatgaatctgcccagggccgagcctggagggtg 71
Db 232 CCTATCCCTATGGTTTCTTCCACAAACATTTCTGCTCTGCTCAGCCTGCTGTGGTG 291
QY 72 ctaccggccgagggccggggaacaatatgaatc 104
Db 292 CTGCCTGTCCCTCAGCCTGAAATAATGAAGTC 324

RESULT 15
US-08-249-554-1
Sequence 1, Application US/08249554
Patent No. 5506209

GENERAL INFORMATION:
APPLICANT: Mukerji, P.
APPLICANT: Seo, A.
APPLICANT: Anderson, S.
APPLICANT: Schaller, J.
TITLE OF INVENTION: Product for the Inhibition of Infection of Mammalian Cells by
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lonnie R. Drayer
ADDRESSEE: ROSS Products Division
ADDRESSEE: Abbott Laboratories
STREET: 625 Cleveland Avenue
CITY: Columbus
STATE: Ohio
COUNTRY: United States
ZIP: 43215
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb storage (B)COMPUTER: Apple Macintosh
OPERATING SYSTEM: Macintosh System 7.1(D)SOFTWARE: Clarisworks 1.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/249,554
FILING DATE: 26-MAY-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA: NO. 5506209 applicable
TELECOMMUNICATION INFORMATION:
TELEPHONE: (614) 624-3774
TELEFAX: (614) 624-3074
TELEX: No. 5506209e
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1065 base pairs
TYPE: Nucleic acid
STRANDEDNESS: Single
TOPOLOGY: Unknown
MOLECULE TYPE: Cloned cDNA representing the product of a human
MOLECULE TYPE: genomic DNA segment.
DESCRIPTION: Human milk 'eta-casein
HYPOTHETICAL:
ANTI-SENSE:
FRAGMENT TYPE:
ORIGINAL SOURCE: Human
ORGANISM: Homo sapiens

STRAIN:
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE: Adult
HAPLOTYPE:
TISSUE TYPE: Mammary Gland
CELL TYPE:
CELL LINE:
ORGANELLE:
IMMEDIATE SOURCE: Human Mammary Gland
LIBRARY:
CLONE:
POSITION IN GENOME:
CHROMOSOME/SEGMENT:
MAP POSITION:
UNITS:
FEATURE:
NAME/KEY:
LOCATION:
IDENTIFICATION METHOD: DNA sequencing and restriction analysis
OTHER INFORMATION: The encoded product of nucleotide SEQ ID NO: 1: is the hum
PUBLICATION INFORMATION:
AUTHORS: B. Lonnerdal et al
TITLE: Cloning and Sequencing of a cDNA encoding human
TITLE: milk beta-casein.
JOURNAL: Federation European Biochemical Society Letters
VOLUME: 269
ISSUE:
PAGES: 153 - 156
DATE: 1990
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO:
US-08-249-554-1

Query Match 7.2%; Score 27.4; DB 1; Length 1065;
Best Local Similarity 55.9%; Pred. No. 2.9;
Matches 52; Conservative 0; Mismatches 41; Indels 0; Gaps 0;
QY 12 cccacgcgtatcggtttacatcttatgaatctgcccagggccgagcctggagggtg 71
Db 211 CCTATCCCTATGGTTTCTTCCACAAACATTTCTGCTCTGCTCAGCCTGCTGTGGTG 270
QY 72 ctaccggccgagggccggggaacaatatgaatc 104
Db 271 CTGCCTGTCCCTCAGCCTGAAATAATGAAGTC 303

Search completed: June 19, 2002, 16:22:15
Job time: 15793 sec

Thu Jun 20 06:56:42 2002

us-09-462-955b-1_copy_611_991.rni

Page 8

Result No.	Score	Query Match	Length	DB	ID	Description
1	35	9.2	524	12	BH175883	BH175883 004_P-17-
2	35	9.2	524	12	C6507IRP	AL612823 T3 end of
3	34.4	9.0	898	9	AL667528	AL667528 AL667528
C 4	34	8.9	1101	12	CNS00DBS	AL067144 Drosophill
5	33.6	8.8	403	10	BG006361	BG006361 QV4 -GN014
6	33.4	8.8	283	9	AO065300	AO065300 TENU2187
C 7	33.2	8.7	1030	9	B6160652	B6160652 B6160652
C 8	33	8.7	644	12	AZ339575	AZ339575 LM0071B12
9	33	8.7	767	9	AV763886	AV763886 AV763886
10	33	8.7	972	12	AQ459250	AQ459250 HS_5078_A
C 11	32.8	8.6	668	12	BH437599	BH437599 BOHAW63TF
12	32.6	8.6	659	9	AA221432	AA221432 My22ell.r
13	32.4	8.5	306	10	D61210	D61210 HUM168A07B
14	32.4	8.5	359	9	A1910101	A1910101 RC-BT236-
15	32.4	8.5	414	12	CNS00POPA	AL084760 Arabidops
16	32.4	8.5	522	12	B30289	B30289 F9H2TF IGF
17	32.4	8.5	537	10	B1011057	B1011057 OV2-FN009

```

/clone_lib="SmbAC1"
/dev_stage="cercariae"
/lab_host="Biomphalaria glabrata"
/note="Vector: pBeloBAC 11; Site_1: Hind III; Partially
Hind III digested and size-selected S. mansoni cercarial
DNA was ligated into Hind III digested pBeloBAC 11 vector
and used to transform E. coli DH10B. The complete library
contains 23808 clones from 4 independent
sizing-ligation-transformations. Average insert size
ranges from 70-127 kb and genome coverage is 7.9-fold."
BASE COUNT      153 a   101 c   126 g   99 t   45 others
ORIGIN

Query Match      9.2%; Score 35; DB 12; Length 524;
Best Local Similarity 50.0%; Pred. No. 1.1;
Matches 69; Conservative 4; Mismatches 65; Indels 0; Gaps 0;

QY 67 aggtgtaccgcgcgaagccggaacaaatgaatcgagttatggcgcccaaat 126
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 59 AGNTACTTCGGATCCGTTGGCCAGACACTATCAGACACAGCTAGTGTGGACWAAACAA 118
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 127 aaagattccatttggaagaacgaatctgttacttctgcttcagtgcaacacact 186
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 119 ACCANATTCAGTTGGAGGAAAAAATCAGAGANNRYRTBCTGGAAGTGATANGACACAC 178
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 187 ttccacgtcacaatcca 204
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 179 TTTCMGGAATCCCCCA 196
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 2
CNS07IRP      524 bp      DNA      linear      GSS 08-OCT-2001
LOCUS      T3 end of clone 004CH09 of library SmbAC1 from strain Puerto-Rican
DEFINITION      of Schistosoma mansoni, genomic survey sequence.
ACCESSION      AL612823
VERSION      AL612823.1 GI:16026029
KEYWORDS      GSS.
SOURCE      Schistosoma mansoni.
ORGANISM      Schistosoma mansoni.
REFERENCE      Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
AUTHORS      Strigoida; Schistosomatidae; Schistosomatidae; Schistosoma.
TITLE      1 (bases 1 to 524)
Le Paslier,M.C., Pierce,R.J., Merlin,F., Hirai,H., Wu,W.,
Williams,D.L., Johnston,D., LoVerde,P.T. and Le Paslier,D.
CONSTRUCTION and characterization of a Schistosoma mansoni
bacterial artificial chromosome library
Genomics 65 (2), 87-94 (2000)
JOURNAL      20247247
MEDLINE
REFERENCE      2 (bases 1 to 524)
Direct Submission
Genoscope.
TITLE      Submitted (05-OCT-2001) Genoscope - Centre National de Sequencage :
JOURNAL      BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT      Partially Hind III digested and size-selected S. mansoni cercarial
DNA was ligated into Hind III digested pBeloBAC 11 vector and used
to transform E. coli DH10B. The complete library contains 23808
clones from 4 independent sizing-ligation-transformations. Average
insert size ranges from 70-127 kb and genome coverage is 7.9-fold.
FEATURES
source
Location/Qualifiers
1..524
/organism="Schistosoma mansoni"
/strain="Puerto-Rican"
/db_xref="taxon:6183"
/clone="004CH09"
/clone_lib="SmbAC1"
/note="end : T3"
BASE COUNT      153 a   101 c   126 g   99 t   45 others
ORIGIN

Query Match      9.2%; Score 35; DB 12; Length 524;
Best Local Similarity 50.0%; Pred. No. 1.1;
Matches 69; Conservative 4; Mismatches 65; Indels 0; Gaps 0;

QY 67 aggtgtaccgcgcgaagccggaacaaatgaatcgagttatggcgcccaaat 126
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 59 AGNTACTTCGGATCCGTTGGCCAGACACTATCAGACACAGCTAGTGTGGACWAAACAA 118
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 127 aaagattccatttggaagaacgaatctgttacttctgcttcagtgcaacacact 186
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 119 ACCANATTCAGTTGGAGGAAAAAATCAGAGANNRYRTBCTGGAAGTGATANGACACAC 178
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 187 ttccacgtcacaatcca 204
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 179 TTTCMGGAATCCCCCA 196
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 3
AL667528      898 bp      mRNA      linear      EST 11-JAN-2002
LOCUS      AL667528 directional larval cDNA library Ciona intestinalis cDNA
DEFINITION      clone 020ZG05 5', mRNA sequence.
ACCESSION      AL667528
VERSION      AL667528.1 GI:18134435
KEYWORDS      EST.
SOURCE      Ciona intestinalis.
ORGANISM      Ciona intestinalis.
REFERENCE      Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
AUTHORS      Phlebobranchia; Clonidae; Ciona.
TITLE      1 (bases 1 to 898)
JOURNAL      Ciona intestinalis directional larval cDNA library
COMMENT      Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
IMPORTANT: this sequence may contain errors. The Ciona intestinalis
library from which the clone was isolated may be contaminated with
cDNAs from bacteria or other Eukarya.
Directional larval cDNA library originate from Dr. M.Branno,
Stazione A. Dohrn, Naples, Italy, and was prepared in
pBluescript2SK+.
FEATURES
source
Location/Qualifiers
1..898
/organism="Ciona intestinalis"
/db_xref="taxon:7719"
/clone="020ZG05"
/clone_lib="directional larval cDNA library"
/note="Vector: pBluescript2SK+"
BASE COUNT      289 a   186 c   157 g   265 t   1 others
ORIGIN

Query Match      9.0%; Score 34.4; DB 9; Length 898;
Best Local Similarity 60.9%; Pred. No. 2.2;
Matches 56; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 169 gcagtgcacgcacacacttccacgtccacaaatccaggtgagtagcttgaggagaagaa 228
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 298 GTAATACATGAACACACATTCACGTCACCAACCGTTGTGAGTATATTGCTTGCAATG 357
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 229 agccgcaagcctctatctaccgtccattgtt 260
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 358 TGCATGAAGCGCTTGGAAACAGTCAGCTGTT 389
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 4
CNS00DBS/c
LOCUS      Drosophila melanogaster genome survey sequence T7 end of BAC #
DEFINITION      BACR27J14 of RPCI-98 library from Drosophila melanogaster (fruit

```


REFERENCE	Mammalia: Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus;
AUTHORS	1 (bases 1 to 644) Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE	Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL	Unpublished (2000)
COMMENT	Contact: Robert B. Weiss University of Utah Genome Center University of Utah Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLCC, UT 84112, USA Tel: 801 585 5606 Fax: 801 585 7177 Email: ddunn@genetics.utah.edu Insert Length: 10000 Std Error: 0.00 Plate: 0071 row: B column: 12 Seq primer: CGTTGTAACAGCAGCGCCAGT Class: plasmid ends High quality sequence stop: 644.

```

FEATURES
source
    .644
    location/Qualifiers
    /organism="Mus musculus"
    /strain="C57BL/6J"
    /db_xref="taxon:10090"
    /clone="UUGCLM0071B12"
    /clone.lib="Mouse 10kb plasmid UUGCLM library"
    /sex="Male"
    /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
    /notes="Vector: PWD42nv; Purified genomic DNA from M.
    musculus C57BL/6J (male) was obtained from the Jackson
    Laboratory Mouse DNA resource
    (http://www.jax.org/resources/documents/dnares/). The DNA
    was hydrodynamically sheared by repeated passage through a
    0.005 inch orifice at constant velocity. The sheared DNA
    was blunt end-repaired with T4 DNA polymerase and T4
    polynucleotide kinase. Adaptor oligonucleotides were
    ligated to the blunt ends in high molar excess. The
    adaptor DNA was purified and size-selected for a 9.5 to
    10.5 kb range using preparative agarose gel
    electrophoresis. Vector DNA was prepared from a derivative
    of PWD42 (g14732114|gb|AF129072.1), a copy-number
    inducible derivative of plasmid R1. The vector was ligated
    with adaptors complementary to the insert adaptors and
    purified. The sheared, adaptor mouse DNA was annealed to
    adaptor vector DNA, and transformed into
    chemically-competent E. coli XL10-Gold (Stratagene) cells
    and selected for ampicillin resistance."
    177 a 186 c 100 g 181 t

```

```

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 762)
Gu,J., Zhao,M., Huang,Q., Xu,X., Li,Y., Peng,Y., Song,H., Xiao,H.,
,L., Xu,S., Gu,W., Qian,B., Liu,F., Qu,J., Gao,X., Cheng,Z., Xu,Z., Zeng
Yang,Y., Gao,G., Zhang,Q., Chen,S., Han,Z. and Chen,Z.
Homo sapiens cDNA MDS clones
Unpublished (2000)
Contact: Zeguag Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzq@chgc.sh.cn
This clone is available at CHGC in Shanghai.
FEATURES             Location/Qualifiers
     source           1..762
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="MDSEEE03"
                     /clone_lib="MDS"
                     /tissue_type="Bone marrow"
                     /cell_type="CD34+ hematopoietic stem/progenitor cell"
                     /lab_host="BM25.8"
                     /notes="Vector: pTriplEx2; Site_1: sfiIA; Site_2: sfiIB"
BASE COUNT      203 a   172 c   177 g   209 t       1 others
ORIGIN

Query Match      8.7%; Score 33; DB 9; Length 762;
Best Local Similarity 48.9%; Pred. No. 6.1;
Matches 87; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 136 catttgataagaacgaatctgtactttgctgtagtcacgcgaaccacctttccacgtc 195
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 471 CAGCTTGTGCGAAGTTACTTGATGTTATTATTTAGATGACTACCACTCTGTCTGATTC 530
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 196 accaatcacggtagtagtctgtggagaagaacgcgaagcctctatatccgtccat 255
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 531 ACCTCGCTGGATTCTAAATGGGTATATTGCCAACTGCGAGCTTTCATTATGGCATTT 590
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 256 ttgttttgcgcgatcgacgcggtgagtgtatctgcgcgcacaaacctctgctaagtc 313
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 591 GTCTTTGAACAATCGTGGTGACATTTGTTTTAACAAAAAAAACTCGGCCCAATC 648
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 10
AQ459250
LOCUS
DEFINITION HS_5078_Al_H03_SP6E RPCI-11 Human Male BAC Library Homo sapiens
genomic clone Plate=654 Col=5 Row=0, DNA sequence.
ACCESSION AQ459250
VERSION AQ459250.1 GI:4637890
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 977)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center

```

University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel.: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu

Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Resear h Genetics (info@resgen.com). BAC end Web Server: <http://www.htsc.washington.edu>

Plate: 654 row: 0 column: 5
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 977.

[illegible]

```

1. .977
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=654 Col=5 Row=0"
/clone_lib="RPC1-11 Human Male BAC Library"
/sex="male"
/note="vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBACE3.6 vector at EcoRI sites"
254 a 223 c 159 q 335 t 6 others

```

	BASE COUNT	ORIGIN
1	0.0000	0.0000
2	0.0000	0.0000
3	0.0000	0.0000
4	0.0000	0.0000
5	0.0000	0.0000
6	0.0000	0.0000
7	0.0000	0.0000
8	0.0000	0.0000
9	0.0000	0.0000
10	0.0000	0.0000
11	0.0000	0.0000
12	0.0000	0.0000
13	0.0000	0.0000
14	0.0000	0.0000
15	0.0000	0.0000
16	0.0000	0.0000
17	0.0000	0.0000
18	0.0000	0.0000
19	0.0000	0.0000
20	0.0000	0.0000
21	0.0000	0.0000
22	0.0000	0.0000
23	0.0000	0.0000
24	0.0000	0.0000
25	0.0000	0.0000
26	0.0000	0.0000
27	0.0000	0.0000
28	0.0000	0.0000
29	0.0000	0.0000
30	0.0000	0.0000
31	0.0000	0.0000
32	0.0000	0.0000
33	0.0000	0.0000
34	0.0000	0.0000
35	0.0000	0.0000
36	0.0000	0.0000
37	0.0000	0.0000
38	0.0000	0.0000
39	0.0000	0.0000
40	0.0000	0.0000
41	0.0000	0.0000
42	0.0000	0.0000
43	0.0000	0.0000
44	0.0000	0.0000
45	0.0000	0.0000
46	0.0000	0.0000
47	0.0000	0.0000
48	0.0000	0.0000
49	0.0000	0.0000
50	0.0000	0.0000
51	0.0000	0.0000
52	0.0000	0.0000
53	0.0000	0.0000
54	0.0000	0.0000
55	0.0000	0.0000
56	0.0000	0.0000
57	0.0000	0.0000
58	0.0000	0.0000
59	0.0000	0.0000
60	0.0000	0.0000
61	0.0000	0.0000
62	0.0000	0.0000
63	0.0000	0.0000
64	0.0000	0.0000
65	0.0000	0.0000
66	0.0000	0.0000
67	0.0000	0.0000
68	0.0000	0.0000
69	0.0000	0.0000
70	0.0000	0.0000
71	0.0000	0.0000
72	0.0000	0.0000
73	0.0000	0.0000
74	0.0000	0.0000
75	0.0000	0.0000
76	0.0000	0.0000
77	0.0000	0.0000
78	0.0000	0.0000
79	0.0000	0.0000
80	0.0000	0.0000
81	0.0000	0.0000
82	0.0000	0.0000
83	0.0000	0.0000
84	0.0000	0.0000
85	0.0000	0.0000
86	0.0000	0.0000
87	0.0000	0.0000
88	0.0000	0.0000
89	0.0000	0.0000
90	0.0000	0.0000
91	0.0000	0.0000
92	0.0000	0.0000
93	0.0000	0.0000
94	0.0000	0.0000
95	0.0000	0.0000
96	0.0000	0.0000
97	0.0000	0.0000
98	0.0000	0.0000
99	0.0000	0.0000
100	0.0000	0.0000

	Query Match	8.7%	Score 33;	DB 12;	Length 977;
	Best Local Similarity	60.7%;	Pred. NO. 6.8;		
	Matches 54;	Conservative 0;	Mismatches 35;	Indels 0;	Gaps 0;
QY	124	aataaaagattccattgggataagaacgaatctgttactttccttcgcagtgcacgcacaacc	183		
Db	708	AATAAGAAATTATAATTTGGGCAATACTTCTTGACCTTGTTCATTTGCACGGGACT	767		
QY	184	actttccacgctcaccaaccaggtagta	212		
Db	768	AGGTATTTCCCTCCTCAAATAGTGGATAATA	796		

RESULT	11	
LOCUS	BH437599/c	
DEFINITION	BH437599 BOHAW63TF BOHA Brassica oleracea genomic clone BOHAW63, DNA 668 bp DNA linear GSS 12-DEC-2001	
ACCESSION	BH437599	
VERSION	BH437599	
KEYWORDS	GSS.	
SOURCE	BH437599.1 GI:17623320	
ORGANISM	Brassica oleracea. Brassica oleracea Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica. 1 (bases 1 to 668)	
REFERENCE	Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M. Whole genome shotgun sequencing of Brassica oleracea	
AUTHORS	Unpublished (2001)	
TITLE	Other GSSs: BOHAW63TR	
JOURNAL	Contact: Chris Town	
COMMENT		

FEATURES
 • source 1. .668
 Class: sheared ends.
 Seq primer: TF
 DNA is from a doubled haploid provided by Tom Osborn.
 Email: cdtownetigr.org

source

```

/organism="Brassica oleracea"
/strain="TOL000DH3"
/db_xref="taxon:3712"
/clone="BOHAW63"
/clone_lib="BOHA"
/note=vector: PHOS1; Site_1: BstXI; 2-3 kb sheared
genomic DNA inserted into PHOS1 using BstXI linkers"
168 a 124 c 184 g 192 t
BASE COUNT
ORIGIN

```

	Query Match	8.6%;	Score 32.8;	DB 12;	Length 668;
	Best Local Similarity	52.1%;	Pred. No. 6.7;		
	Matches	73;	Conservative	0;	Mismatches 67;
					Indels 0;
					Gaps 0;
Qy	123	caataaagattccatttgataaagacgaatctgttactttgcttcagtgacgcgcaac	182		
Db	644	CTACAAGAGCTGCCACACAGTAGGCACAGTTTGTTCCTTGTCTACCGTGCACAGAC	585		
Qy	183	cactttccacgtcaccatccaggtgtagctgtctgtggagaagaacgcgcaagcctct	242		
Db	584	GACCTTTGTCAACATCACCCCATCTAAAGACGCTTCCTCTGAGGACAGTTACTAGAACTCG	525		
Qy	243	atctaccgtccattgtttt	262		
Db	524	AGCTTCCAACCATGACTTCT	505		

RESULT 12

AA221432	AA221432	659 bp	linear	EST 13-FEB-1997
LOCUS	my22ell.r1	Barstead mouse pooled organs	MPLRB4 Mus musculus cDNA	
DEFINITION	clone IMAGE:696620 5' similar to gb:L23769	Mouse		
	microfibril-associated glycoprotein (MOUSE);	RNA sequence.		
ACCESSION	AA221432			
VERSION	AA221432.1	GI:1840686		
KEYWORDS	EST.			
SOURCE	house mouse.			

REFERENCE
AUTHORS

TITLE
 The WashU-HHMI Mouse EST Project
COMMENT
 Unpublished (1996)
 Contact: Marra M/Mouse EST Project
 WashU-HHMI Mouse EST Project
 Washington University School of MedicineP
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LLNL ; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:430180
 Seg primer: -28m13 rev2 ET from Amersham
 High quality sequence stop: 472.

FEATURES	SOURCE
----------	--------

```

i. .039
source
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone_image:696620"
/clone_lib="Barstead mouse pooled organs MPLRB4"
/sex="mixed"
/tissue_type="pooled organs"
/dev_stage="7 day"
/lab_host="DH10B"
/notes="Organ: pooled; Vector: pT7m3D-Pac (Pharma

```


VERSION AL084760.1 GI:5285900
 KEYWORDS GSS.
 SOURCE thale cress.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 REFERENCE 1 (bases 1 to 414)
 AUTHORS Salanoubat,M., Choisme,N., Artiguenave,F., Brottier,P., Wincker,P.,
 Samson,D., Saurin,W., Weissenbach,J. and Quetier,F.
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 414)
 Genoscope.
 Direct Submission
 TITLE Submitted (25-JUN-1999) Genoscope - Centre National de Sequencage :
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
 - Web : www.genoscope.cns.fr)
 FEATURES Location/Qualifiers
 source
 1..414
 /organism="Arabidopsis thaliana"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /clone_lib="IGF"
 /clone="F9H2"
 /note="end : T7"
 BASE COUNT 153 a 66 c 58 g 137 t
 ORIGIN

Query Match 8.5%; Score 32.4; DB 12; Length 414;
 Best Local Similarity 60.0%; Pred. No. 7.3;
 Matches 54; Conservative 0; Mismatches 36; Indels 0; Gaps 0;
 QY 95 aatatgaatcgagttatggcgggccacataaaagattccatttgataagaacgaat 154
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 25 AATATGAATAAAGTTTGTATACACTTGAATTAATAATATCCAAAAGCGCTGTTAACTAAT 84
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 155 ctgttacttgcgtgcagtcgcaacca 184
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 85 CTCTTCTATATAATAACACAGTACTAAACCA 114

Search completed: June 19, 2002, 14:04:28
 Job time: 7536 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2002, 15:48:13 ; Search time 5904.86 Seconds
(without alignments)
995.850 Million cell updates/sec

Title: US-09-462-955B-1_COPY_711_991

Perfect score: 281

Sequence: 1 aatcgagttatggcgggcc.....ccccaacctctgtaacccc 281

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb.ba.*

2: gb.htg.*

3: gb.in.*

4: gb.om.*

5: gb.ov.*

6: gb.pat.*

7: gb.ph.*

8: gb.pl.*

9: gb.pr.*

10: gb.ro.*

11: gb.sts.*

12: gb.sy.*

13: gb.un.*

14: gb.vi.*

15: em.ba.*

16: em.fun.*

17: em.hum.*

18: em.in.*

19: em.mu.*

20: em.om.*

21: em.or.*

22: em.ov.*

23: em.pat.*

24: em.ph.*

25: em.pl.*

26: em.ro.*

27: em.sts.*

28: em.un.*

29: em.vi.*

30: em.htg.hum.*

31: em.htg.inv.*

32: em.htg.other.*

33: em.htgo.inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description

1	191	68.0	1291	6	AR172770	AR172770 Sequence
2	191	68.0	1291	14	CFDCG	M29963 Coconut fol
3	37	13.2	210115	9	AL442127	AL442127 Human DNA
4	36.4	13.0	173848	9	AC090043	AC090043 Homo sapi
5	33	11.7	1438	14	IAU86144	U86144 Influenza A
6	32.8	11.7	29176	8	AP001312	AP001312 Arabidops
7	32.8	11.7	39435	3	CEG49F8	Z70206 Caenorhabdi
8	32.4	11.5	3002	1	AF269928	AF269928 Staphyloc
9	32.4	11.5	3002	6	AX145246	AX145246 Sequence
10	32.4	11.5	3295	1	AF269634	AF269634 Staphyloc
11	32.4	11.5	3295	6	AX144954	AX144954 Sequence
12	32.4	11.5	3518	1	AF270074	AF270074 Staphyloc
13	32.4	11.5	3518	6	AX145392	AX145392 Sequence
14	32.4	11.5	3985	1	AF269816	AF269816 Staphyloc
15	32.4	11.5	3985	6	AX145134	AX145134 Sequence
16	32.4	11.5	105805	2	AP003745	AP003745 Oryza sat
17	32.2	11.5	14907	1	AE007142	AE007142 Mycobacte
18	32.2	11.5	58280	1	MTV014	AL021646 Mycobacte
19	32.2	11.5	168835	9	HS652L8	AL031734 Human DNA
20	32	11.4	245282	2	AC093320	AC093320 Mus muscu
21	31.8	11.3	183539	2	AC091354	AC091354 Rattus no
22	31.6	11.2	118994	9	AC010634	AC010634 Homo sapi
23	31.6	11.2	141456	2	AC108092	AC108092 Homo sapi
24	31.6	11.2	159160	2	AC018624	AC018624 Homo sapi
25	31.6	11.2	172516	2	AC096154	AC096154 Rattus no
26	31.6	11.2	176094	9	AC106799	AC106799 Homo sapi
27	31.4	11.2	98637	2	AL365359	AL365359 Homo sapi
28	31.4	11.2	117919	2	AC053470	AC053470 Mus muscu
29	31.4	11.2	127168	9	HS1163J1	AL031588 Human DNA
30	31.4	11.2	143454	2	AC026820	AC026820 Homo sapi
31	31.4	11.2	316704	9	AF239258S3	AF223391 Homo sapi
32	31.2	11.1	1991	8	AF462831	AF462831 Arabidops
33	31.2	11.1	87246	9	AC002539	AC002539 Homo sapi
34	31.2	11.1	95643	8	F6N15	AC069299 Arabidops
35	31.2	11.1	167234	8	AC004708	AC004708 CIC5B11.1
36	31.2	11.1	175644	2	AC106743	AC106743 Homo sapi
37	31.2	11.1	197119	8	ATCHR1V1	AL161471 Arabidops
38	31	11.0	31	6	AR172771	AR172771 Sequence
39	31	11.0	75052	2	AC090558	AC090558 Homo sapi
40	31	11.0	166682	9	AC068194	AC068194 Homo sapi
41	31	11.0	167178	2	AC025672	AC025672 Homo sapi
42	31	11.0	169193	2	AC068523	AC068523 Homo sapi
43	31	11.0	172182	2	AC079816	AC079816 Mus muscu
44	31	11.0	174365	2	AL626785	AL626785 Mus muscu
45	31	11.0	174783	2	AC015565	AC015565 Homo sapi

ALIGNMENTS

RESULT 1	AR172770	AR172770	1291 bp	DNA	linear	PAT 17-DEC-2001
LOCUS	Sequence 1 from patent US 6303345.					
DEFINITION	AR172770					
ACCESSION	AR172770.1	GI:17912261				
VERSION	Unknown.					
KEYWORDS	Unclassified.					
SOURCE	Unclassified.					
ORGANISM	1 (bases 1 to 1291)					
REFERENCE	Rohde,W., Becker,D., Randles,J.W., Hehn,A. and Salamini,F.					
AUTHORS	Use of a virus DNA as promoter					
TITLE	Patent: US 6303345-A 1 16-OCT-2001;					
JOURNAL	Location/Qualifiers					
FEATURES	source					
	1..1291					
	/organism="unknown"					
BASE COUNT	336 a	323 c	332 g	300 t		
ORIGIN						

Query Match 68.0%; Score 191; DB 6; Length 1291;
Best Local Similarity 100.0%; Pred. No. 5.5e-51;

```

Matches 191; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 aatcgagttatggcgcccaataaagaattccatttggataagaacgaatctgtta 60
|||||
Db 1101 AATCGAGTTATGGCGGCCCAATAAAGATTTCATTGGATAGACGAATCTGTTA 1160

Qy 61 ctttgcgttgagtcgcaacacacatttccacgtcccaatccagtgagtgagctgtg 120
|||||
Db 1161 CTTTGTCTGCAGTCGACGAACACACTTCCACGTCACCAATCAGGTAGTAGTGTGCTG 1220

Qy 121 gagaagaaacccgaagctctatctaccgtccattgttttttgcgcgacgagctgtg 180
|||||
Db 1221 GAGAAGAAACCCGACGCTCTATCTACCGTCCATTGTTTTTGGCGATCGACGGCTG 1280

Qy 181 agttgatctgg 191
|||||
Db 1281 AGTTGATCTGG 1291

RESULT 2
CFDGC
LOCUS          1291 bp ss-DNA      circular VRL 02-AUG-1993
DEFINITION    Coconut foliar decay virus, complete genome.
ACCESSION     M29963
VERSION       M29963.1 GI:323306
KEYWORDS      circular; complete genome.
SOURCE        Coconut foliar decay virus DNA, clones NMORG, pORG, and pORGSAU3A.
ORGANISM      Viruses; ssDNA viruses; Nanovirus.

REFERENCE
AUTHORS       Rohde,W., Randles,J.W., Langridge,P. and Hanold,D.
TITLE         Nucleotide sequence of a circular single-stranded DNA associated
              with coconut foliar decay virus
JOURNAL        Virology 176, 648-651 (1990)
MEDLINE        90266484
COMMENT       Draft entry and printed sequence for [1] kindly submitted by
              W.Rhode, 15-MAR-1989, for release after publication.
FEATURES
source        1..1291
              /organism="Coconut foliar decay virus"
              /db_xref="taxon:12474"
              40..70
              /note="stem-loop structure"
              103..975
              /note="ORF 1"
              /codon_start=1
              /db_xref="GI:323307"
              /protein_id="AAA42894.1"
              /translation="MGSSIRRCFNLNVEEERANVVRRIESLNLVYAIYGVDEVAPS
              TQRLHQGFILHUKTRRLQGLKVLGNDRIHLEPTRGSDQNDRDYCSKERVLLLEHGV
              PRPKRRLAQRFAEEPELRLDPGGYRRCVHGASVETWRAAENPPFPYHNWQ
              LEVLSAIGEPADDRITLWICGRDGDGKSVFAKYLGLKPDWFTYTCGTRKDVLYQIE
              DPKRLILIDVPRCNLEYLNLYALLECVKNRAFSSDKYEPLSYLGFHDHVLVFNVLDP
              YLKISRDRIKLWNI"
              314..775
              /note="ORF 2"
              /codon_start=1
              /protein_id="AAA42895.1"
              /db_xref="GI:323308"
              /translation="MTGFTWSPVPTNRIETTVRRNGCFSTESRLVLESKGDWPN
              DLLRLMSAKTQADTADYALTELRWNGQGPLKIRSHFHTITGSLKCLRSRSQT
              IAQSSYADETETGSPCLPNISDSSPTGSHVVEPTCTSTSTRYNEI"
              complement(422..568)
              /note="ORF 6"
              /codon_start=1
              /protein_id="AAA42896.1"
              /db_xref="GI:323309"
              /translation="MEMGTDFQRPILSTPPKLRVQIRFGIRLGLPGGVHVPQIVGP
              IVAF"
              639..797
              /note="ORF 3"
              /codon_start=1
              /protein_id="AAA42897.1"

```

```

/db_xref="GI:323310"
/translation="MTRRRRREVRCVQISRTQARLVLHMWNQKGRIVPHRGPKTK
FNPRCTQV"
complement(823..987)
/note="ORF 5"
/codon_start=1
/protein_id="AAA42898.1"
/db_xref="GI:323311"
/translation="MTHTLNIPQFYSVPADFQIIRDIGKYEYMHMVEPKITKGFV
RTECPVLNTP"
1098..1286
/note="ORF 4"
/codon_start=1
/protein_id="AAA42899.1"
/db_xref="GI:323312"
/translation="MNRVMGGPTIKDSIWIRTNLLCLCQTOPLSTSPIQVSLLEKK
AASLYLPSICFAIGRLS"

BASE COUNT    336 a 323 c 332 g 300 t
ORIGIN

Query Match      68.0%; Score 191; DB 14; Length 1291;
Best Local Similarity 100.0%; Pred. No. 5.5e-51;
Matches 191; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 aatcgagttatggcgcccaataaagaattccatttggataagaacgaatctgtta 60
|||||
Db 1101 AATCGAGTTATGGCGGCCCAATAAAGATTTCATTGGATAGACGAATCTGTTA 1160

Qy 61 ctttgcgttgagtcgcaacacacatttccacgtcccaatccagtgagtgagctgtg 120
|||||
Db 1161 CTTTGTCTGCAGTCGACGAACACACTTTCACGTCACCAATCAGGTAGTAGTGTGCTG 1220

Qy 121 gagaagaaacccgaagctctatctaccgtccattgttttttgcgcgacgagctgtg 180
|||||
Db 1221 GAGAAGAAACCCGACGCTCTATCTACCGTCCATTGTTTTTGGCGATCGACGGCTG 1280

Qy 181 agttgatctgg 191
|||||
Db 1281 AGTTGATCTGG 1291

RESULT 3
AL442127
LOCUS          Human DNA sequence from clone RP11-297I6 on chromosome 13, complete
DEFINITION     sequence.
ACCESSION      AL442127
VERSION        AL442127.7 GI:12584450
KEYWORDS       HTG.
SOURCE         Homo sapiens
ORGANISM       Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
              1 (bases 1 to 210115)
              Tromans,A.
              Direct Submission
              Submitted (26-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
              CB10 1SA, UK. E-mail enquiries: humqueresanger.ac.uk
              requests: clonerequest@sanger.ac.uk
              On Jan 28, 2001 this sequence version replaced gi:12578234.
              During sequence assembly data is compared from overlapping clones.
              Where differences are found these are annotated as variations
              together with a note of the overlapping clone name. Note that the
              variation annotation may not be found in the sequence submission
              corresponding to the overlapping clone, as we submit sequences with
              only a small overlap as described above.
              This sequence has been finished according to sequence map criteria
              as follows. An attempt is made to resolve all sequencing problems,
              such as compressions and repeats, but not necessarily within known
              annotated repeat sequence elements. Where the sequence is
              ambiguous, there is an annotation using the 'unsure' feature key.
              The following abbreviations are used to associate primary accession

```


numbers given in the feature table with their source databases:
 Em:, EMBL: SW:, SWISSPROT: Tr:, TREMBL: Wp:, WORMPEP; Information
 on the WORMPEP database can be found at
<http://www.sanger.ac.uk/projects/C.elegans/wormpep> This sequence
 was generated from part of bacterial clone contigs of human
 chromosome 13, constructed by the Sanger Centre Chromosome 13
 Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr13>
 RP11-29716 is from the library RPCI-11.2 constructed by the group
 of Pieter de Jong. For further details see
<http://www.chori.org/bacpac/home.htm>
 VECTOR: pBACe3.6
 This sequence is the entire insert of clone RP11-29716 The true
 left end of clone RP11-207D10 is at 130589 in this sequence. The
 true right end of clone RP11-272L14 is at 88495 in this sequence.

FEATURES

source	Location/Qualifiers
misc_feature	1..210115 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="13" /clone="RP11-29716" /clone_lib="RPCI-11.2" 131..1822 /note="CpG island" /evidence=not_experimental 2357..2504 /note="37 copies 4 mer atat 75% conserved" 2363..2502 /note="70 copies 2 mer at 77% conserved" 3453..3755 /note="AluY repeat: matches 1..303 of consensus" 5180..5450 /note="AluSc repeat: matches 1..270 of consensus" 5939..5978 /note="10 copies 4 mer tgga 100% conserved" 9204..9248 /note="15 copies 3 mer gtt 75% conserved" 10198..10397 /note="MER20 repeat: matches 7..218 of consensus" 12023..12404 /note="MSTRD repeat: matches 1..394 of consensus" 13891..14196 /note="AluY repeat: matches 1..302 of consensus" 15743..16019 /note="MLT1D repeat: matches 190..505 of consensus" 16034..16726 /note="MLT1-INTERNAL repeat: matches 42..880 of consensus" 16727..17162 /note="MLT1D repeat: matches 40..482 of consensus" 19077..19259 /note="L2 repeat: matches 2335..2504 of consensus" 19751..20050 /note="AluSg repeat: matches 1..302 of consensus" 20222..20521 /note="AluSg repeat: matches 1..297 of consensus" 25657..25831 /note="Charlie5 repeat: matches 2373..2585 of consensus" 26571..26606 /note="9 copies 4 mer acac 100% conserved" 27856..28166 /note="AluY repeat: matches 1..311 of consensus" 28333..28578 /note="MIR repeat: matches 7..260 of consensus" 28668..28789 /note="AluJb repeat: matches 1..124 of consensus" 29414..29579 /note="MIR repeat: matches 66..259 of consensus" 29712..29855 /note="L2 repeat: matches 1508..1650 of consensus" 30170..30368 /note="L2 repeat: matches 2056..2275 of consensus" 30667..30744 /note="L2 repeat: matches 2672..2749 of consensus"
repeat_region	32012..32087 /note="L1ME2 repeat: matches 6058..6131 of consensus" 32363..33462 /note="CpG island" /evidence=not_experimental 32558..32703 /note="73 copies 2 mer cc 63% conserved" 32623..32754 /note="Sequence from overlapping clone bA272L14 (AL138689): Assembly confirmed by restriction digest." 33621..33682 /note="MIR repeat: matches 85..148 of consensus" 33747..33968 /note="MIR repeat: matches 15..257 of consensus" 34151..34515 /note="L1M4 repeat: matches 4316..4702 of consensus" 34520..34746 /note="MLT1A repeat: matches 143..365 of consensus" 34756..34890 /note="MLT1A repeat: matches 3..138 of consensus" 34891..35118 /note="MLT1A repeat: matches 162..365 of consensus" 35316..35546 /note="MLT1A repeat: matches 133..365 of consensus" 35547..35633 /note="MLT1A repeat: matches 63..138 of consensus" 35900..36016 /note="MLT1A repeat: matches 1..117 of consensus" 36404..36604 /note="MLT1C repeat: matches 264..466 of consensus" 36715..37003 /note="AluSg1 repeat: matches 1..289 of consensus" 37056..37085 /note="15 copies 2 mer aa 93% conserved" 37616..37989 /note="THE1C repeat: matches 1..371 of consensus" 39265..39720 /note="MER92B repeat: matches 147..626 of consensus" 39732..40157 /note="THE1C repeat: matches 1..426 of consensus" 40158..40295 /note="MER92B repeat: matches 1..148 of consensus" 40338..42198 /note="L1MCB repeat: matches 115..2032 of consensus" 42660..42698 /note="MERSA repeat: matches 146..187 of consensus" 42996..43434 /note="MLT1C repeat: matches 1..497 of consensus" 44101..44184 /note="MLT1H repeat: matches 448..531 of consensus" 44362..44400 /note="MLT1G repeat: matches 231..269 of consensus" 44516..44571 /note="MLT1H repeat: matches 72..127 of consensus" 45161..45458 /note="AluSg repeat: matches 1..298 of consensus" 45700..45897 /note="MTR repeat: matches 28..236 of consensus" 46028..46270 /note="MER45C repeat: matches 1..228 of consensus" 46466..46695 /note="MER45C repeat: matches 714..952 of consensus" 47322..47737 /note="MSTC repeat: matches 1..405 of consensus" 48351..48562 /note="MLT1J repeat: matches 300..497 of consensus" 48579..49021 /note="L1MC4 repeat: matches 7494..7944 of consensus" 49719..50097 /note="THE1C repeat: matches 3..371 of consensus" 50098..51657 /note="THE1C-internal repeat: matches 1..1580 of consensus"


```

/product="CLC-d chloride channel; anion channel protein"
/protein_id="BAB01934.1"
/db_xref="GI:9294082"
/translation="MVEDNLNGNSNYNGEGDPESNTLNLOPLVKNRKLSTPLA
LVKARVSHIESLDEIENNDLFKIDWRKSKAQVLYVFLKWLACLVGLTGLIATL
INLAENAVIKYLLAVHFTQERYVTGLMVLVGNLGLTLVSLVLCVFCRPAAGPG
IPEIKAVLYGVDTFMFGATTMIKIVGSGIAVAGADLDKEGPLVHGSCIASLLGQ
GTDNHRKRWLRVFNDRDRDLITCGSAGAAFCVCAAFSPGVGLFALBEVATWRS
ALLWRTFSTAVVHLLKHFETIEICNSKGLFGKGLIMDFDVSHVTYVHTYDIPVM
LIGVIGLSLVNHLKHLVRLYLNEKGIHKVLLSLVSLFTSVCLYGLPELAK
CKPCDSIDEICPTNGRSGNFKQHCPCGYNDLTLTNDVARNLFSSWTPNEE
GMGSLIFFVILGLFTGFIATPSGLFPIILMGAAYGRMLGAAAGSTSIDQGLI
AVLGAALMAGSMRTVSLVIFLETLNLLLPITMIVLLIARTGDSFNPSYDII
LHLKGLPLEANPEWRNLTVELGDAKPPVTLQGVKSVNIVDLKNTTNAFV
LDEAPVGLATGATELHGLLRAHLVKVLRKWFLEKRRTEEWVRKFPWDELA
ERENFDVAITSMEYVDLPLTNTPTVTVMENSVAKALVFRQVGLRHLLIYP
KIQASGMVPGVILTRDRLAYNIQAPFLLEKSKGKTH"
complement(join(20497..20704,20817..20950,21039..21380,
21519..21632,21759..21842,21918..21997,22092..22146,
22231..22298,22378..22422,22498..22627,22716..22830,
22944..23104))
/note="embI|CAB12631.1
gene_id:MYF5.5
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB01935.1"
/db_xref="GI:9294083"
/translation="MNSSLSFGGRFPIINCRVSAASAPAFVASVPLPSTSDSLA
CSLQPHFQSCGGTQENLRHRAVDASDFKRYGVDDFTFDSCLRWRCKRAKLA
VGGSDNALIGLYQEVHTVVDIPECKNFCDAHPNINAAIELREGIKVDFVFPF
DEDOGTGLRVQMAVTHSTNLRAPEYKNGKVQVSLVWNRNERNKADQLST
YLWRKGGNSKFLIHSVWVNFOTSTNIFGNWRHLLGERDFWEHVGIDILSDLS
SEGOATRAFSLDLKHLKHYVFGSSVADLVAGVIGLSLATSRCRQEORLFTSVL
NLKRYMFLMKRLTQYFLKSDPSVSKIEVKEARLSFEKTIQRLPNSLCSISWH
HADAVNPLSLTIGSDVVVDPDRGLDLSRQMLESPVSEKRMRSQSSQSLNAKE
EKRPWLRAKELSLIQAGNKQTSERNNTLQRLIISCGWESFKEVLENTGICOCRRF
RSF"
complement(24625..24880)
/note="gene_id:MYF5.6
unknown protein"
/codon_start=1
/pseudo
/evidence=not_experimental
BASE COUNT 9717 a 5015 c 4729 g 9715 t
ORIGIN

```

```

Query Match 11.7%; Score 32.8; DB 8; Length 29176;
Best Local Similarity 52.1%; Pred. No. 17;
Matches 73; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 24 aataaagatttcatttgataaagcaaatctgttactttgtctgagtcagcagcaacc 83
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 23766 AATCAACAATTCCTTTTGGCAAGCATAGATAGAAATTTTGTTCACACTTTGTATCT 23707
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 84 actttccacgtcccaatcaggtgagtagcttgctggagagaagacgcagcctcta 143
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 23706 TGTAACTATTTTACCITTCCTCTTTGATTACACAGTTTGAGAGGAACAAGAGACTATA 23647
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 144 tctacgttcatttggttttt 163
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 23646 TATGCAATTTTTCCTTTTTCCTTTTTCCTTTTTCCTTTTTCCTTTTTCCTTTT 23627
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 7
CEC49F8/c CEC49F8 39435 bp DNA linear INV 24-JAN-2002
LOCUS Caenorhabditis elegans cosmid C49F8, complete sequence.
DEFINITION
ACCESSION 270206
VERSION 270206.1 GI:1229047
KEYWORDS HFG; monocarboxylate transporter like.
SOURCE Caenorhabditis elegans.

```

ORGANISM

Caenorhabditis elegans

Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.

REFERENCE

1 (sites)

AUTHORS

none.

TITLE

Genome sequence of the nematode C. elegans: a platform for

JOURNAL

Investigating biology. The C. elegans Sequencing Consortium

MEDLINE

Science 282 (5396), 2012-2018 (1998)

REMARK

99069613

REFERENCE

2 (bases 1 to 39435)

AUTHORS

Hemby, C.

TITLE

Direct Submission

JOURNAL

Submitted (16-MAR-1996) Nematode Sequencing Project, Sanger

COMMENT

Institute, Hinxton, Cambridge CB10 1SA, England and Department of

JOURNAL

Genetics, Washington University, St. Louis, MO 63110, USA. E-mail:

COMMENT

jes@sanger.ac.uk or rw@nematoe.wustl.edu

COMMENT

Coding sequences below are predicted from computer analysis, using

COMMENT

predictions from GeneFinder (P. Green, U. Washington), and other

COMMENT

available information.

COMMENT

Current sequence finishing criteria for the C. elegans genome

COMMENT

sequencing consortium are that all bases are either sequenced

COMMENT

unambiguously on both strands, or on a single strand with both a

COMMENT

dye primer and dye terminator reaction, from distinct subclones.

COMMENT

Exceptions are indicated by an explicit note.

COMMENT

IMPORTANT: This sequence is not the entire insert of clone C49F8.

COMMENT

It may be shorter because we only sequence overlapping sections

COMMENT

once, or longer because we arrange for a small overlap between

COMMENT

neighbouring submissions.

COMMENT

The true left end of clone C49F8 is at 1 in this sequence. The true

COMMENT

left end of clone F02C12 is at 39332 in this sequence. The true

COMMENT

right end of clone F46G10 is at 1490 in this sequence. The start of

COMMENT

this sequence (1..1490) overlaps with the end of sequence 250177.

COMMENT

The end of this sequence (39332..39435) overlaps with the start of

COMMENT

sequence 254269.

COMMENT

For a graphical representation of this sequence and its analysis

COMMENT

see: - http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?

COMMENT

name=C49F8

COMMENT

IMPORTANT: This sequence is NOT necessarily the entire insert of

COMMENT

the specified clone. It may be shorter because we only sequence

COMMENT

overlapping sections once, or longer because we arrange for a small

COMMENT

overlap between neighbouring submissions.

FEATURES

Location/Qualifiers

source

1..39435

/organism="Caenorhabditis elegans"

/db_xref="taxon:6239"

/chromosome="X"

/clone="C49F8"

complement(join(1006..1201,1257..1401,1454..1727,

1779..1871))

/gene="F46G10.6"

complement(join(1006..1201,1257..1401,1454..1727,

1779..1871))

/gene="F46G10.6"

/note="contains similarity to pfam domain: PF00010

(Helix-loop-helix DNA-binding domain), Score=63.6,

E-value=1.4e-15, N=1"

/codon_start=1

/protein_id="CAA94125.1"

/db_xref="GI:3875036"

/db_xref="SPTREMBL:O18711"

/translation="MSAIVEDVLLSSMKMKQFRKRHHSDSDSSSPKSPKSPSM

DDRRHNELEERRRRHHKIDHETILKDAIPLLDGKSSPALILKRAVEPIHYMOTKL

SSQKATLEDLTRNKLLEERLRESGSSSSRLPALAVSSQMQLTPILIPQMONT

AQLSQYQQANITIAQSTNPAQLDGLIALNDAILLGLSFQSPSLDSDAPGTFPSG

FYPCAFSPVDQOMAVKI"

complement(join(12393..12614,12767..12955,13353..13460,

13504..13556,14826..14866,15183..15266,15849..16375))

/gene="C49F8.1"

complement(join(12393..12614,12767..12955,13353..13460,

13504..13556,14826..14866,15183..15266,15849..16375))

/gene="C49F8.1"

/note="predicted using GeneFinder"

```

/codon_start=1
/protein_id="CAA94124.1"
/db_xref="GI:3875035"
/db_xref="SPTREMBL:Q18710"
/translation="MLNKTQNLINQRHCSOLKSLFTTABFFQFTSDSLFTNFPKAKLI
YSTRVCVGHMFSELSFSSFFHFSSFFPFPFLFPLFHLFAYTHPFFSRNN
FSPSPFIFFSFCNCPYIFILIFHASSVREHHIMSSQRSMFGEKAVIEMDIIDPFSVNL
DTPADLKVPPTTYGPMLSLTQLATPTMKRQDONOTHRDDEKKNNAETAGEPSAKKRS
VSYLAFVAAPPTSNQOQYTOAQOYPAQOEIKPSOTVPOLSAHHGVYVGPQQAQPOL
QYPSPTAAPAPPEQPTLNNPLEVKQSAQVLFNDPDRQMLARLSQSTGSGSVSNP
LSARDIPCCSDGRSRTIHLRLSPESLSVQIARQAQFNPDKWLRLADMKLSALGT
PMHSTHYAARSIGL"
gene complement(join(23105..23177,24031..24187,24238..24671,
24759..25318,25734..26165,26214..26347,27098..27317,
27370..27597,27777..27881,28376..28456))
/genes="C49f8.2"
CDS complement(join(23105..23177,24031..24187,24238..24671,
24759..25318,25734..26165,26214..26347,27098..27317,
27370..27597,27777..27881,28376..28456))
/genes="C49f8.2"
/notes="Similarity to Hamster monocarboxylate transporter
(SR-MOT1_CRLIO), contains similarity to Pfam domain:
PF01587 (Monocarboxylate transporter), Score=722.1,
E-value=8.1e-214, N=1
cDNA EST yk14d9.3 comes from this gene
cDNA EST yk294c8.3 comes from this gene
cDNA EST yk294c8.5 comes from this gene
cDNA EST yk219b11.5 comes from this gene"
/codon_start=1
/protein_id="CAA94126.1"
/db_xref="GI:3875037"
/db_xref="SPTREMBL:Q18712"
/translation="MSAKSNQKFIYFIFHPDKIYLLVERLETPESEMARRRPS
MTRKMGVSCWENHQPFSPSDEPRVRFAPAPVARSDDSDAESVETSLRPPV
KDGIGWLVVLSFLHVCADGSCFVIFVQIHFQCFRFSVMTASMLSPLI
MSPVAGIVLIGCRMSIIIGASICTGCIIMFCSHIFFMFLGCGVGMSTIYN
AAIVITVYFEKRGATSPAGSCTGVTVIYIPLNLSMIYASVSDIRILIFPA
VMYVLAIGVFIDKVEDSDKDFKRFQYIAKMDDEKENNTPTDAFMRHATSL
PNIILNHHSGSIQSTCDAAKARKSDKPTRSKSVLFONKHOMSSIPEYMLNTLA
NLHLDLEANSCTVYRAQRRVVISKVSMSVDQINEDEAFHINLFSSTESDD
ESSDSEMSNDGSSSELSEKNLEPTTKLINKLENVAFRATSSVPSARMNRNS
VAAGTSMNFGVRVIAASQATSHATNLIAAGKIPSAPLVARKNRKSMGKPIQGFYK
KWLTEKPIQVEYIHCRAYIYIALSVLCYFLIDLPVYCVFYESTIDLKWSEANYI
YYSIGTSNFTSVLFTLADTKRHITTVIYASSMGVGITIMFTRINLSLVELVIGV
CFGVTTISNVLOSILVTCIFEDVNLFOVAYSLIAMLEGVASLIGPPIFALYREATGG
YTVYFISGIFALASGFFGFEMFYEMNKNRTGDKDSIESGKPTELTNGNROTSGG
AEQETLLDV"
gene complement(join(37306..37596,38318..38467,38535..38624))
/genes="C49f8.3"
CDS complement(join(37306..37596,38318..38467,38535..38624))
/genes="C49f8.3"
/notes="predicted using Genefinder
contains similarity to Pfam domain: PF01682 (DB module),
Score=199.7, E-value=1.4e-56, N=1
cDNA EST yk196b6.5 comes from this gene"
/codon_start=1
/protein_id="CAA94123.1"
/db_xref="GI:3875034"
/db_xref="SPTREMBL:Q18709"
/translation="MMRLTSCLLLAATYATYGAENGIDFDPNGNGTGAASNNKPCIE
RKVADQVFGSCCRFPVPECRSLCIYETNAIEARVYIMHAIQPSRCLRYKYLPAIVHC
AAOTHNSCECRSGNGVAGELGECQVQCQPSNPRRFRWVKSLRKDMVVCVLARWDQIMQ
CHOSGLRARKVRPVNA"
BASE COUNT 13447 a 6457 c 7151 g 12380 t
ORIGIN

```

```

Query Match 11.7%; Score 32.8; DB 3; Length 39435;
Best Local Similarity 51.4%; Pred. No. 17;
Matches 76; Conservative 0; Mismatches 72; Indels 0; Gaps 0;
Qy 94 tcaccaatccaggtagcttgcctgagagaagccgcaagcctctctaccgtcc 153
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 21277 TCACCTAICTTGTGTCGTCTTCAAAAGCATAGATCGTAGAACATATTTTCGTCA 21218
Qy 154 atttgTTTTgCGatcgagcggtgagttgatctcgccgcaaaaacccctctgctaagtc 213
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 21217 ACTTGTTTTTTTAAATGGAATATCAAAATGCTTTGTCTACACAACAATTCGAAAAATTC 21158
Qy 214 cctgctgaagtaataatgacgcgcgggg 241
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 21157 ACGTTTCATAAAAAACAATTTTGGGG 21130
RESULT 8
AF269928 3002 bp DNA linear BCT 01-AUG-2000
LOCUS Staphylococcus epidermidis strain SRI clone step.1032f01 genomic
DEFINITION sequence.
ACCESSION AF269928
VERSION AF269928.1 GI:9623829
KEYWORDS
SOURCE Staphylococcus epidermidis.
ORGANISM Staphylococcus epidermidis
Bacteria; Firmicutes; Bacillus/Clostridium group;
Bacillus/Staphylococcus group; Staphylococcus.
REFERENCE 1 (bases 1 to 3002)
AUTHORS Kimmerly,W.J., Taylor,J.David., Nelsen,A.J., Godlevski,M.M.,
Rubino,M.A., Nelson,F.J., Rivers,P.R., Tortuella-Miller,I.,
Listenbee,S., Ashanti,C., Altshuller,G., Mamo,L., Shepherd,N.S.,
Fuchs,R., Fleming,T., Guan,X., Du,L., Cain,D.H., Miller,G.S. and
Furdon,P.J.
TITLE Transposon-mediated sequencing of the Staphylococcus epidermidis
genome
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3002)
AUTHORS Taylor,J.David., Kimmerly,W.J., Nelsen,A.J., Godlevski,M.M.,
Rubino,M.A., Nelson,F.J., Rivers,P.R., Tortuella-Miller,I.,
Listenbee,S., Ashanti,C., Altshuller,G., Mamo,L., Shepherd,N.S.,
Fuchs,R., Fleming,T., Guan,X., Du,L., Cain,D.H., Miller,G.S. and
Furdon,P.J.
TITLE Direct Submission
JOURNAL Submitted (22-MAY-2000) Departments of Genomic Sciences and
Bioinformatics, Genetics Directorate, Glaxo Wellcome, Inc., 5 Moore
Drive, Research Triangle Park, North Carolina 27709-3398, USA
FEATURES
source
1..3002
/organism="Staphylococcus epidermidis"
/strain="SRI"
/db_xref="taxon:1282"
/clone="step.1032f01"
BASE COUNT 899 a 595 c 418 g 1090 t
ORIGIN
Query Match 11.5%; Score 32.4; DB 1; Length 3002;
Best Local Similarity 53.1%; Pred. No. 20;
Matches 69; Conservative 0; Mismatches 61; Indels 0; Gaps 0;
Qy 100 atccagtgatgtagcttgcgagagaagccgcaagcctctctaccgtccattgt 159
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 627 ATTTAAGTGATTATTAGTAGATAAGAAATGCCTGAGACAAATGATTCGCCAGCATTTTA 686
Qy 160 ttttgCGatcgagcggtgagttgatctcgccgcaaaaacccctctgtaagtcgcgtgc 219
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 687 TTTATTGGCAATWAGAACCGAGTTGAACCTTGGCTTAATCTCAATTTTCAACTCTAGTCC 746
Qy 220 taagtataaa 229
||||| ||| |||
Db 747 TCATTACCAA 756
RESULT 9
AX145246 3002 bp DNA linear PAT 31-MAY-2001
LOCUS AX145246
DEFINITION Sequence 3968 from Patent WO0134809.

```

```

ACCESSION AX145246
VERSION AX145246.1 GI:14283811
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
    source
        1..3002
        /organism="synthetic construct"
        /db_xref="taxon:32630"
        /note="synthetic nucleic acid sequence"
BASE COUNT 899 a 595 c 418 g 1090 t
ORIGIN
Query Match 11.5%; Score 32.4; DB 6; Length 3002;
Best Local Similarity 53.1%; Pred. No. 20;
Matches 69; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 100 atccaggtgagtagctgtcgagagaagccgcaagcctctatctaccgtccattgt 159
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 627 ATTTAAGTGATTATCTAGTATAGAAAATGCTGAGACAATGATTTGCCAGGCATTTTA 686
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 160 ttitgcgcatcgagcggtgagttgattctgagcgcacaaacctctgtaagtcctgc 219
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 687 TTTATTGGCAATTGACACCGAGTTGAACCTGGGCTAAATCTCAATTTCAACTCTAGTCC 746
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 220 taagtataaa 229
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 747 TCATTACCAA 756

RESULT 10
AF269634 3295 bp DNA linear BCT 01-AUG-2000
LOCUS Staphylococcus epidermidis strain SRI clone step.1014g06 genomic
DEFINITION
ACCESSION AF269634
VERSION AF269634.1 GI:9623530
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
    source
        1..3295
        /organism="synthetic construct"
        /db_xref="taxon:32630"
        /note="synthetic nucleic acid sequence"
BASE COUNT 971 a 667 c 456 g 1201 t
ORIGIN
Query Match 11.5%; Score 32.4; DB 6; Length 3295;
Best Local Similarity 53.1%; Pred. No. 20;
Matches 69; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 100 atccaggtgagtagctgtcgagagaagccgcaagcctctatctaccgtccattgt 159
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 776 ATTTAAGTGATTATCTAGTATAGAAAATGCTGAGACAATGATTTGCCAGGCATTTTA 835
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 160 ttitgcgcatcgagcggtgagttgattctgagcgcacaaacctctgtaagtcctgc 219
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 836 TTTATTGGCAATTGACACCGAGTTGAACCTGGGCTAAATCTCAATTTCAACTCTAGTCC 895
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 220 taagtataaa 229
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 896 TCATTACCAA 905

RESULT 12
AF270074 3518 bp DNA linear BCT 01-AUG-2000
LOCUS Staphylococcus epidermidis strain SRI clone step.1047b09 genomic
DEFINITION
ACCESSION AF270074
VERSION AF270074.1 GI:9623979
KEYWORDS
SOURCE
ORGANISM
    Staphylococcus epidermidis.
    Staphylococcus epidermidis.

```

```

Bacteria; Firmicutes; Bacillus/Clostridium group;
Bacillus/Staphylococcus group; Staphylococcus.
REFERENCE
AUTHORS
1 (bases 1 to 3518)
Kimberly, W.J., Taylor, J. David., Nelsen, A.J., Godlevski, M.M.,
Rubino, M.A., Nelson, F.J., Rivers, P.R., Torruella-Miller, I.,
Listenbee, S., Ashanti, C., Altshuller, G., Mamo, L., Shepherd, N.S.,
Fuchs, R., Fleming, T., Guan, X., Du, L., Cain, D.H., Miller, G.S. and
Furdon, P.J.
TITLE
Transposon-mediated sequencing of the Staphylococcus epidermidis
genome
JOURNAL
Unpublished
REFERENCE
AUTHORS
2 (bases 1 to 3518)
Taylor, J. David., Kimberly, W.J., Nelsen, A.J., Godlevski, M.M.,
Rubino, M.A., Nelson, F.J., Rivers, P.R., Torruella-Miller, I.,
Listenbee, S., Ashanti, C., Altshuller, G., Mamo, L., Shepherd, N.S.,
Fuchs, R., Fleming, T., Guan, X., Du, L., Cain, D.H., Miller, G.S. and
Furdon, P.J.
TITLE
Direct Submission
JOURNAL
Submitted (22-MAY-2000) Departments of Genomic Sciences and
Bioinformatics, Genetics Directorate, Glaxo Wellcome, Inc., 5 Moore
Drive, Research Triangle Park, North Carolina 27709-3398, USA
FEATURES
LOCATION/Qualifiers
SOURCE
1. .3518
/organism="Staphylococcus epidermidis"
/strain="SRI"
/db_xref="taxon:1282"
/clone="step.1047b09"
BASE COUNT 1091 a 677 c 495 g 1255 t
ORIGIN
Query Match 11.5%; Score 32.4; DB 1; Length 3518;
Best Local Similarity 53.1%; Pred. No. 20;
Matches 69; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

Qy 100 atccaggtgagtagctgtggaagaagccgcaagcctctatcaccgtccattgt 159
Db 2590 ATTTAAGTGATTCTTAGTATAGAAAATGCTGAGACAATGATTTGCCAGGCATTTA 2649

Qy 160 ttttgcgcgatcgacgcgtgagttgatctggcgccaaaacccctctgctaagtcctgc 219
Db 2650 TTTATTGGCAATTGACAAACCGAGTGAACCTGGGCTAATCTCAATTTTCAACTCTAGTCC 2709

Qy 220 taagtataaa 229
Db 2710 TCATTACCAA 2719

RESULT 13
AX145392
LOCUS AX145392 3518 bp DNA linear PAT 31-MAY-2001
DEFINITION Sequence 4114 from Patent WO0134809.
ACCESSION AX145392
VERSION AX145392.1 GI:14283957
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct
artificial sequence.
REFERENCE
AUTHORS 1 (bases 1 to 3518)
Kimberly, W.J.
TITLE Staphylococcus epidermidis nucleic acids and proteins
JOURNAL Patent: WO 0134809-A 4114 17-MAY-2001;
GLAXO GROUP LIMITED (GB)
LOCATION/Qualifiers
FEATURES
SOURCE
1. .3518
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="synthetic nucleic acid sequence"
BASE COUNT 1091 a 677 c 495 g 1255 t
ORIGIN
Query Match 11.5%; Score 32.4; DB 6; Length 3518;

```

```

Best Local Similarity 53.1%; Pred. No. 20;
Matches 69; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

Qy 100 atccaggtgagtagctgtggaagaagccgcaagcctctatcaccgtccattgt 159
Db 2590 ATTTAAGTGATTCTTAGTATAGAAAATGCTGAGACAATGATTTGCCAGGCATTTA 2649

Qy 160 ttttgcgcgatcgacgcgtgagttgatctggcgccaaaacccctctgctaagtcctgc 219
Db 2650 TTTATTGGCAATTGACAAACCGAGTGAACCTGGGCTAATCTCAATTTTCAACTCTAGTCC 2709

Qy 220 taagtataaa 229
Db 2710 TCATTACCAA 2719

RESULT 14
AF269816
LOCUS AF269816 3985 bp DNA linear BCT 01-AUG-2000
DEFINITION Staphylococcus epidermidis strain SRI clone step.1024b10 genomic
sequence.
ACCESSION AF269816
VERSION AF269816.1 GI:9623714
KEYWORDS
SOURCE Staphylococcus epidermidis.
ORGANISM Staphylococcus epidermidis
Bacteria; Firmicutes; Bacillus/Clostridium group;
Bacillus/Staphylococcus group; Staphylococcus.
REFERENCE
AUTHORS 1 (bases 1 to 3985)
Kimberly, W.J., Taylor, J. David., Nelsen, A.J., Godlevski, M.M.,
Rubino, M.A., Nelson, F.J., Rivers, P.R., Torruella-Miller, I.,
Listenbee, S., Ashanti, C., Altshuller, G., Mamo, L., Shepherd, N.S.,
Fuchs, R., Fleming, T., Guan, X., Du, L., Cain, D.H., Miller, G.S. and
Furdon, P.J.
TITLE
Transposon-mediated sequencing of the Staphylococcus epidermidis
genome
JOURNAL Unpublished
REFERENCE
AUTHORS 2 (bases 1 to 3985)
Taylor, J. David., Kimberly, W.J., Nelsen, A.J., Godlevski, M.M.,
Rubino, M.A., Nelson, F.J., Rivers, P.R., Torruella-Miller, I.,
Listenbee, S., Ashanti, C., Altshuller, G., Mamo, L., Shepherd, N.S.,
Fuchs, R., Fleming, T., Guan, X., Du, L., Cain, D.H., Miller, G.S. and
Furdon, P.J.
TITLE
Direct Submission
JOURNAL Submitted (22-MAY-2000) Departments of Genomic Sciences and
Bioinformatics, Genetics Directorate, Glaxo Wellcome, Inc., 5 Moore
Drive, Research Triangle Park, North Carolina 27709-3398, USA
FEATURES
LOCATION/Qualifiers
SOURCE
1. .3985
/organism="Staphylococcus epidermidis"
/strain="SRI"
/db_xref="taxon:1282"
/clone="step.1024b10"
BASE COUNT 1175 a 779 c 583 g 1448 t
ORIGIN
Query Match 11.5%; Score 32.4; DB 1; Length 3985;
Best Local Similarity 53.1%; Pred. No. 20;
Matches 69; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

Qy 100 atccaggtgagtagctgtggaagaagccgcaagcctctatcaccgtccattgt 159
Db 381 ATTTAAGTGATTCTTAGTATAGAAAATGCTGAGACAATGATTTGCCAGGCATTTA 440

Qy 160 ttttgcgcgatcgacgcgtgagttgatctggcgccaaaacccctctgctaagtcctgc 219
Db 441 TTTATTGGCAATTGACAAACCGAGTGAACCTGGGCTAATCTCAATTTTCAACTCTAGTCC 500

Qy 220 taagtataaa 229
Db 501 TCATTACCAA 510

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2002, 17:05:17 ; Search time 3798.65 Seconds
(without alignments)
127.006 Million cell updates/sec

Title: US-09-462-955B-1_COPY_711_991
Perfect score: 281
Sequence: 1 aatgagtgatggcgccgccc.....cccccaactctgtaacccc 281

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802.*
1: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1980.DAT.*
2: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1981.DAT.*
3: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1982.DAT.*
4: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1983.DAT.*
5: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1984.DAT.*
6: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1985.DAT.*
7: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1986.DAT.*
8: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1987.DAT.*
9: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1988.DAT.*
10: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1989.DAT.*
11: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1990.DAT.*
12: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1991.DAT.*
13: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1992.DAT.*
14: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1993.DAT.*
15: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1994.DAT.*
16: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1995.DAT.*
17: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1996.DAT.*
18: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1997.DAT.*
19: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1998.DAT.*
20: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1999.DAT.*
21: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA2000.DAT.*
22: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA2001A.DAT.*
23: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA2001B.DAT.*
24: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	32.4	11.5	3002	AAH54604	S. epidermidis gen
2	32.4	11.5	3295	AAH54312	S. epidermidis gen
3	32.4	11.5	3518	AAH54750	S. epidermidis gen
4	32.4	11.5	3985	AAH54492	S. epidermidis gen
5	32.2	11.5	4411529	AA199682	Mycobacterium tube
6	31.4	11.2	460	AA196825	Human polynucleoti
7	31	11.0	31	AAQ56057	Coconut Foliar Dec
8	31	11.0	31	AAQ56057	Gemini virus DNA f
9	31	11.0	31	AAQ56057	Gemini virus DNA f

c	10	30.4	10.8	974	21	AA666416	Human secreted pro
c	11	30	10.7	265	14	AAQ47710	TGF-beta-like clon
c	12	30	10.7	2088	21	AAZ45827	Degenerate DNA enc
c	13	30	10.7	2088	22	AA511992	Human degenerate D
c	14	30	10.7	2272	17	AA111104	Transforming growth
c	15	30	10.7	2272	22	AA744421	Human TGF-beta MPl
c	16	29.8	10.6	1636	22	AAH77087	Human hepatic nucl
c	17	29.6	10.5	30600	20	AAZ32023	Human METHI relate
c	18	29.6	10.5	30601	22	AAZ32023	259361 cDNA clone.
c	19	29.4	10.5	400	18	AAV78089	Staphylococcus aur
c	20	29.4	10.5	1194	23	AA54726	Staphylococcus aur
c	21	29.4	10.5	14078	18	AAV74502	Staphylococcus aur
c	22	29.4	10.5	48452	23	ABL07108	Drosophila melanog
c	23	29.2	10.4	8877	23	AA57138	DNA encoding Droso
c	24	29.2	10.4	8877	23	ABL13920	Drosophila melanog
c	25	28.6	10.2	315	22	ABA18519	Human nervous syst
c	26	28.6	10.2	591	23	AA580984	DNA encoding novel
c	27	28.6	10.2	3031	22	ABA18518	Human nervous syst
c	28	28.6	10.2	5531	22	AAK78665	Human immune/haema
c	29	28.2	10.0	858	22	AA123310	Human breast cance
c	30	28.2	10.0	1414	19	AAV26248	Genomic DNA SEQ ID
c	31	28.2	10.0	1497	21	AAZ50194	Male fusion plasm
c	32	28	10.0	978	23	AA582110	DNA encoding novel
c	33	28	10.0	978	23	AA586806	DNA encoding novel
c	34	28	10.0	1509	23	AA52691	E. coli DNA for ce
c	35	28	10.0	2493	23	ABL26334	Drosophila melanog
c	36	28	10.0	3737	21	AA63739	Maize RNA-directed
c	37	28	10.0	5164	23	ABL01909	Drosophila melanog
c	38	28	10.0	10953	22	AA53689	Genomic sequence #
c	39	28	10.0	10953	22	AAK90032	Human digestive sy
c	40	28	10.0	30598	23	ABL01908	Drosophila melanog
c	41	28	10.0	80450	21	AAZ2295	BAC containing rep
c	42	27.8	9.9	522	18	AAV75313	Staphylococcus aur
c	43	27.6	9.8	1317	23	ABL13049	Drosophila melanog
c	44	27.6	9.8	1378	14	AAQ42310	Rgt precursor. Or
c	45	27.6	9.8	3317	23	ABL13048	Drosophila melanog

ALIGNMENTS

RESULT 1
AAH54604
ID AAH54604 standard; DNA; 3002 BP.
AC AAH54604;
XX
XX
DT 03-SEP-2001 (first entry)
XX
XX
DE S. epidermidis genomic polynucleotide sequence SEQ ID NO:3968.
XX
XX
KW Staphylococcus epidermidis SR1 strain; infection; diagnosis;
XX vaccination; endocarditis; ds.
XX
XX
OS Staphylococcus epidermidis.
XX
XX
PN WO200134809-A2.
XX
XX
PD 17-MAY-2001.
XX
XX
PF 09-NOV-2000; 2000WO-US30782.
XX
XX
PR 09-NOV-1999; 99US-0164258.
XX
XX
PA (GLAXO) GLAXO GROUP LTD.
XX
XX
PI Kimmerly WJ;
XX
XX
DR WPI; 2001-316495/33.
XX
XX
PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
XX useful for vaccinating against infections, e.g. endocarditis -

WPI; 2001-316495/33.

Nucleic acids encoding polypeptides from *Staphylococcus epidermidis*, useful for vaccinating against infections, e.g. endocarditis -

Claim 8; Page 1500-1502; 2188pp; English.

AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides (II), given in AAG81454 to AAG83120, from *Staphylococcus epidermidis*. (I) and (II) can have antibacterial activity and therefore can be used in vaccination. The nucleic acids (I) may be used to produce the *S. epidermidis* polypeptides (II) via the production of vectors containing them which are used to produce hosts cells which express the polypeptides. The polypeptides (II) (and/or nucleic acids) may then be used to vaccinate subjects and to raise antibodies against the bacteria. The polypeptides may also be used to assay for other inhibitors of their activity and therefore identify compounds that may be used for the treatment of *S. epidermidis* infections, e.g. endocarditis. AAH53971 to AAH55090 represent specifically claimed *S. epidermidis* genomic DNA polynucleotide sequences from the present invention. AAH55091 to AAH55098 represent oligonucleotide sequences and primers which are used in the exemplification of the present invention.

N.B. The present invention specifically claims all the polynucleotide sequences given in the sequence listing of the present specification, however the sequence listing only goes up to SEQ ID NO:4454 so even though sequences are given in the disclosure for SEQ ID NO:4465 to 4472, no sequences are present for SEQ ID NO:4455 to 4464.

Sequence 3985 BP; 1175 A; 779 C; 583 G; 1448 T; 0 other;

[illegible]

RESULT	4
AAH54492	
ID	AAH54492 standard; DNA; 3985 BP.
XX	
XX	AAH54492;
XX	
XX	03-SEP-2001 (first entry)
DT	
XX	S. epidermidis genomic polynucleotide sequence SEQ ID NO:3856.
DE	
XX	Staphylococcus epidermidis SRI strain; infection; diagnosis;
XX	KW vaccination; endocarditis; ds.
KW	
XX	
XX	Staphylococcus epidermidis.
OS	
XX	WO200134809-A2.
PN	
PN	17-MAY-2001.
PD	
XX	09-NOV-2000; 2000WO-US30782.
XX	
PF	
XX	09-NOV-1999; 99US-0164258.
XX	
PR	(GLAX) GLAXO GROUP LTD.
XX	
PA	
XX	
XX	Kimmerly WJ;
PI	

[illegible][illegible]

XX Nucleic acids encoding polypeptides from *Staphylococcus epidermidis*,
PT useful for vaccinating against infections, e.g. endocarditis -
XX
XX
XX Claim 8; Page 1808-1809; 2188pp; English.
XX
XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC (I), given in AAH81454 to AAH83120, from *Staphylococcus epidermidis*.
CC (II) and (II) can have antibacterial activity and therefore can be used
CC in vaccination. The nucleic acids (I) may be used to produce the
CC *S. epidermidis* polypeptides (II) via the production of vectors
CC containing them which are used to produce host cells which express the
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC used to vaccinate subjects and to raise antibodies against the bacteria.
CC The polypeptides may also be used to assay for other inhibitors of their
CC activity and therefore identify compounds that may be used for the
CC treatment of *S. epidermidis* infections, e.g. endocarditis. AAH53971 to
CC AAH55090 represent specifically claimed *S. epidermidis* genomic DNA
CC polynucleotide sequences from the present invention. AAH55091 to
CC polynucleotide sequences represent oligonucleotide sequences and primers which are used
CC in the exemplification of the present invention.
CC N.B. The present invention specifically claims all the polynucleotide
CC sequences given in the sequence listing of the present specification,
CC however the sequence listing only goes up to SEQ ID NO:4454 so even
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
CC no sequences are present for SEQ ID NO:4455 to 4464.
XX
XX Sequence 3518 BP; 1091 A; 677 C; 495 G; 1255 T; 0 other:
XX

XX Fleischmann RD, White OR, Fraser CM, Venter JC;
 XX WPI; 2001-647261/74.
 XX
 XX Evaluating strain variation of Mycobacterium tuberculosis, comprises
 XX determining the nucleotide sequence of the strain at positions in the
 XX genome corresponding to positions where M. tuberculosis strains CDC
 XX 1551 and H37Rv differ
 XX
 XX Claim 3; SEQ ID NO 1; 3pp + Sequence Listing; English.
 XX
 XX The invention relates to evaluating strain variation within and between
 XX different populations of the tuberculosis bacterial pathogen,
 XX Mycobacterium tuberculosis or related Mycobacterium by determining the
 XX nucleotide sequence of the first strain at positions in the complete
 XX sequence of the genome that correspond to positions that differ in the
 XX nucleotide sequences of M. tuberculosis strains CDC 1551 (AAI99683) and
 XX H37Rv (AAI99682). The method is useful for evaluating strain variation of
 XX M. tuberculosis and has valuable application in the fields of
 XX tuberculosis genetics, epidemiology, patient treatment and epidemic
 XX monitoring.
 XX Note: The sequence data for this patent did not form part of the printed
 XX specification, but was obtained in electronic format directly from USPTO
 XX at seqdata.uspto.gov/sequence.html?docID=6294328B1.
 XX
 XX Sequence 4411529 BP; 758565 A; 1449983 C; 1444602 G; 758379 T; 0 other;
 XX
 XX
 XX Query Match 11.5%; Score 32.2; DB 22; Length 4411529;
 XX Best Local Similarity 57.4%; Pred. No. 15;
 XX Matches 58; Conservative 0; Mismatches 43; Indels 0; Gaps 0;
 XX
 XX QY 108 gagtagctgtggaagaagcgcgaagcctctatctatctacgtccattgtttgagc 167
 XX Db 3577728 GAGCGGCACGGCGGAACAAACACCGCGAGCTGCCGTCTGCGTGGATCGGTGAC 3577669
 XX
 XX QY 168 gatcggagcgtgagttgatctggcgcaaaacacctgtct 208
 XX Db 3577668 GCACTGTTGCCGAGCGGCAGAGGAGATCCAACTGCT 3577628
 XX
 XX RESULT 6
 XX AAI86825/c
 XX ID AAI86825 standard; cDNA; 460 BP.
 XX AC AAI86825;
 XX
 XX DT 06-NOV-2001 (first entry)
 XX DE Human polynucleotide SEQ ID NO 6885.
 XX
 XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 XX tissue growth factor; immunomodulatory; cancer; leukaemia;
 XX nervous system disorders; arthritis; inflammation; ss.
 XX OS Homo sapiens.
 XX
 XX PN WO200164835-A2.
 XX XX
 XX PD 07-SEP-2001.
 XX
 XX PF 26-FEB-2001; 2001WO-US04927.
 XX
 XX PR 28-FEB-2000; 2000US-0515126.
 XX PR 18-MAY-2000; 2000US-0577409.
 XX
 XX PA (HYSE-) HYSEQ INC.
 XX
 XX PI Tang YT, Liu C, Drmanac RT;
 XX WPI; 2001-514838/56.
 XX DR

DR P-PSDB; AAO06894.
 XX
 XX Isolated nucleic acids and polypeptides, useful for preventing
 XX diagnosing and treating e.g. leukaemia, inflammation and immune
 XX disorders -
 XX
 XX PS Claim 1; SEQ ID NO 6885; 1399pp + Sequence Listing; English.
 XX
 XX The invention relates to human polynucleotides (AAI79941-AAI93841) and
 XX the encoded proteins (AAO00010-AAO13910) that exhibit activity relating to
 XX cytokine, cell proliferation or cell differentiation or which may induce
 XX production of other cytokines in other cell populations. The
 XX polynucleotides and polypeptides are useful in gene therapy, vaccines or
 XX peptide therapy. The polypeptides have various cytokine-like activities,
 XX e.g. stem cell growth factor activity, haematopoiesis regulating
 XX activity, tissue growth factor activity, immunomodulatory activity and
 XX activin/inhibin activity and may be useful in the diagnosis and/or
 XX treatment of cancer, leukaemia, nervous system disorders, arthritis and
 XX inflammation.
 XX Note: The sequence data for this patent did not form part of the printed
 XX specification, but was obtained in electronic format directly from WIPO
 XX at ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX SQ Sequence 460 BP; 114 A; 115 C; 141 G; 88 T; 2 other;

Query Match 11.2%; Score 31.4; DB 22; Length 460;
 Best Local Similarity 48.1%; Pred. No. 0.55;
 Matches 89; Conservative 0; Mismatches 96; Indels 0; Gaps 0;
 QY 85 ctctccacgtccaccacacgaagtagctgtagctgtagaagaagcgcgaagcctctat 144
 Db 377 CTGTATCATCTGCCAAGCTGCTCTCAAGCTTTGGCATCAAGCAATCTCGCCCTCAGC 318
 QY 145 ctacgtccattgttttgcgcgacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 204
 Db 317 CTCCCATCATCTGGGATTACAGGTGCACATAGCCCTATAAGGGGTTTGGCTTCCAGTTC 258
 QY 205 tgctaaagtcgcgtctaagataataatagcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 264
 Db 257 TGACTTTGAGGAGGTCATTGGAAACAGACCCCTGGGCGCTGCTTCCCGCTGAGCCCCAC 198
 QY 265 caacc 269
 Db 197 TGCCC 193

RESULT 7
 AAO56057/c
 ID AAO56057 standard; DNA; 31 BP.
 XX
 XX AAO56057;
 XX AC
 XX DT 12-AUG-1994 (first entry)
 XX DE Coconut Foliar Decay Virus promoter stem-loop.
 XX KW Coconut Foliar Decay Virus; CFDV; strong promoter; tissue-specific;
 XX phloem-specific; stem-loop structure; transgenic plant; ds.
 XX OS Coconut Foliar Decay Virus.
 XX
 XX FH Key Location/Qualifiers
 XX stem_loop 1..31
 XX /tag= a
 XX /function= promoter
 XX /note= "loop has homology to geminivirus sequence"
 XX
 XX PN DE4306832-C.
 XX
 XX PD 24-FEB-1994.
 XX
 XX PF 04-MAR-1993; 93DE-4306832.

```
XX PR 04-MAR-1993; 93DE-4306832.
XX PA (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX PI Becker D, Randles JW, Rohde W, Salamini F;
XX PP WPI; 1994-058406/08.
XX DR Use of coconut foliar decay virus DNA as promoter - for
XX PT tissue-specific gene expression in transgenic plants
XX PS Claim 1; Fig 2; 8pp; German.
XX A DNA fragment from the CFDV genome can be used as a phloem-specific
XX CC promoter in the construction of transgenic plants. The promoter is
XX CC strong; it has 30-50% of the activity of the CamV 35S promoter in
XX CC tobacco plants.
XX SQ Sequence 31 BP; 5 A; 11 C; 11 G; 4 T; 0 other;

Query Match 11.0%; Score 31; DB 15; Length 31;
Best Local Similarity 100.0%; Pred. No. 0.23;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 231 agccgcggggctagtagtattaccccgcggt 261
DB 31 AGCCGCGGGGCTAGTATTACCCCGCGGCT 1

RESULT 8
AAAX02688/c
ID AAX02688 standard; DNA; 31 BP.
XX AC AAX02688;
XX DT 10-MAY-1999 (first entry)
XX DE Gemini virus DNA fragment stem loop.
XX KW Stem loop; coconut foliar decay virus; CFDV; bacterial promoter;
XX KW yeast promoter; tissue-specific; phloem-specific; CamV 35S promoter; ss.
XX OS Gemini virus.
XX FH Key Location/Qualifiers
FT stem_loop 1..31 /*tag= a
FT misc_binding 1..11 /*tag= b
FT misc_binding 21..31 /*tag= c
FT /*note= "Region binds to nucleotides 21 to 31"
FT /*note= "Region binds to nucleotides 1 to 11"
XX DE19730502-A1.
XX PN 21-JAN-1999.
XX PD 16-JUL-1997; 97DE-1030502.
XX PF 16-JUL-1997; 97DE-1030502.
XX PR (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX PA Becker D, Hehn A, Randles JW, Rohde W, Salamini F;
XX PI WPI; 1999-096863/09.
XX DR Coconut foliar decay virus promoters - for gene expression in
XX PT bacteria and yeasts
XX PS Disclosure; Fig 2; 12pp; German.
XX This invention describes a coconut foliar decay virus (CFDV) DNA fragment
XX CC that includes the stem-loop structure of CFDV DNA but lacks the
XX CC translation start codons of open reading frames ORF1 and/or ORF2. The
XX CC new CFDV DNA fragment is useful as a bacterial or yeast promoter, as a
XX CC promoter for tissue-specific (especially phloem-specific) gene expression
XX CC in plants and for production of chimeric constructs for transient or
XX CC stable expression. Certain fragments of CFDV DNA have stronger promoter
XX CC activity in E. coli than the CamV 35S promoter.
XX SQ Sequence 31 BP; 5 A; 11 C; 11 G; 4 T; 0 other;

Query Match 11.0%; Score 31; DB 20; Length 31;
Best Local Similarity 100.0%; Pred. No. 0.23;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 231 agccgcggggctagtagtattaccccgcggt 261
DB 31 AGCCGCGGGGCTAGTATTACCCCGCGGCT 1

RESULT 9
AAAX02686/c
ID AAX02686 standard; DNA; 31 BP.
XX AC AAX02686;
XX DT 10-MAY-1999 (first entry)
XX DE Gemini virus DNA fragment stem loop.
XX KW Stem loop; coconut foliar decay virus; CFDV; bacterial promoter;
XX KW yeast promoter; tissue-specific; phloem-specific; CamV 35S promoter; ss.
XX OS Gemini virus.
XX FH Key Location/Qualifiers
FT stem_loop 1..31 /*tag= a
FT misc_binding 1..11 /*tag= b
FT misc_binding 21..31 /*tag= c
FT /*note= "Region binds to nucleotides 21 to 31"
FT /*note= "Region binds to nucleotides 1 to 11"
XX DE19730535-A1.
XX PN 21-JAN-1999.
XX PD 16-JUL-1997; 97DE-1030535.
XX PF 16-JUL-1997; 97DE-1030535.
XX PR (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX PA Becker D, Hehn A, Randles JW, Rohde W, Salamini F;
XX PI WPI; 1999-096867/09.
XX DR Coconut foliar decay virus promoters - for gene expression in
XX PT plants, bacteria and yeasts
XX PS Disclosure; Fig 2; 12pp; German.
XX This invention describes a coconut foliar decay virus (CFDV) DNA fragment
XX CC that includes the stem-loop structure of CFDV DNA but lacks the
XX CC translation start codons of open reading frames ORF1 and/or ORF2. The
XX CC new CFDV DNA fragment is useful as a bacterial or yeast promoter, as a
XX CC promoter for tissue-specific (especially phloem-specific) gene expression
XX CC in plants and for production of chimeric constructs for transient or
```

CC stable expression. Certain fragments of CFV DNA have stronger promoter
 CC activity in E. coli than the CaMV 35S promoter.
 XX
 SQ Sequence 31 BP; 5 A; 11 C; 11 G; 4 T; 0 other;

Query Match 11.0%; Score 31; DB 20; Length 31;
 Best Local Similarity 100.0%; Pred. No. 0.23;
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 231 agccgcggggctagattaccccccgcgcct 261
 |||||
 DB 31 AGCCGCGGGGCTAGTATTACCCCGCGCT 1

RESULT 10
 AAC66416/c
 ID AAC66416 standard; DNA; 974 BP.

XX AAC66416;

AC AAC66416;

XX 14-FEB-2001 (first entry)

DT Human secreted protein coding sequence SEQ ID NO: 17.

DE Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
 KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
 KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;
 KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; human; secreted protein; ss.

XX Homo sapiens.

OS WO200058350-A1.

XX 05-OCT-2000.

XX 22-MAR-2000; 2000WO-US07483.

XX 26-MAR-1999; 99US-0126596.

XX 22-DEC-1999; 99US-0171552.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM, Komatsoulis G;

XX WPI: 2000-602357/57.

XX P-PSDB; AAB32008.

XX Nucleic acid molecules encoding human secreted proteins, used in
 PT preventing, treating or ameliorating a disorder, e.g. Alzheimer's and
 PT Parkinson's diseases and cancers -

XX Claim 1; Page 337; 423pp; English.

XX The invention relates to the isolation of genes AAC66410-C66458 encoding
 CC 49 human secreted proteins AAB32002-B32050. The genes can be used to
 CC generate fusion proteins by linking to the gene for the human
 CC immunoglobulin G Fc portion (SEQID) for increasing the stability of
 CC the fusion protein as compared to the human protein only. The genes and
 CC proteins are useful for preventing, ameliorating or treating medical
 CC conditions, e.g. by protein or gene therapy. The genes are isolated
 CC from a range of human tissues disclosed in the specification. The
 CC nucleic acids, proteins, antibodies and (ant)agonists are useful in
 CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
 CC and ovarian cancer, and other cancers of the adrenal gland, bone, bone
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemia; (d)
 CC wound healing; (e) neurological diseases e.g. cerebral anoxia and
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal

CC and parasitic infections.

XX

SQ Sequence 974 BP; 229 A; 252 C; 263 G; 230 T; 0 other;

Query Match 10.8%; Score 30.4; DB 21; Length 974;
 Best Local Similarity 54.5%; Pred. No. 1.7;
 Matches 61; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 138 cctctatctacgcgtccattgttttgcgcgatcgacggctgagttgatctggcgcaa 197
 |||||
 DB 441 CCTCTAAGATCCCTTCAACTCCGAATCTCTGATTTGACTGCGAGGAAACAAAGCCCTAG 382

QY 198 aaacctctgtcgaagtcctgctaaagtataaaatagccgcgggctagtatt 249
 |||||
 DB 381 AAACATGGGTTACAGCTGGGCTAGGCTTGCTTAGCGGCCAGGCGAGTTT 330

RESULT 11

AAQ47710

ID AAQ47710 standard; cDNA to mRNA; 265 BP.

XX AAQ47710;

AC AAQ47710;

XX 11-FEB-1994 (first entry)

DT TGF-beta-like clone MP-121.

DE Human; transforming growth factor; beta; TGF-beta; pharmaceutical;
 KW bone; cartilage; tooth; wound repair; immunosuppressor;
 KW organ transplant; cosmetic surgery; antibody; diagnosis; ss.

XX Homo sapiens.

OS

XX Key

XX Location/Qualifiers

XX CDS

XX 2..265

XX /*tag= a

XX WO9316099-A.

XX 19-AUG-1993.

XX 12-FEB-1993; 93WO-EP00350.

XX 12-FEB-1992; 92EP-0102324.

XX (BIOP-) BIOPHARM GES BIOTECHNOLOGISCHEN ENTWICKL.

XX Hoetten G, Neidhardt H;

XX WPI: 1993-272824/34.

XX P-PSDB; AAR45447.

XX New transforming growth factor-beta family proteins and DNA -
 PT used in tissue and wound repair, in treatment of bone, cartilage
 PT and tooth defects, and antibodies for diagnosis

XX Claim 4; Page 18; 29pp; English.

XX The sequences given in AAQ47709-10 represent embryo and liver derived
 CC human transforming growth factor-beta (TGF-beta) genes respectively.
 CC The proteins encoded by these sequences may be used in a pharmaceutical
 CC composition for the treatment of various bone, cartilage or tooth
 CC defects and in tissue and wound repair processes. These proteins
 CC may also be used as immunosuppressors in organ transplants and in
 CC cosmetic surgery. Antibodies raised against these proteins may be
 CC used for diagnostic purposes.

XX Sequence 265 BP; 58 A; 80 C; 69 G; 58 T; 0 other;

Query Match

Best Local Similarity 10.7%; Score 30; DB 14; Length 265;

59.3%; Pred. No. 1.4;

	Matches	51;	Conservative	0;	Mismatches	35;	Indels	0;	Gaps	0;
QY	192	cgccaaaaacctctgactgaagtcocggtcaagtataataatgccgcgggggctagttattac	251							
Db	19	cgcattgaactctctgactgaagcaggtgccactacacatagcaggtgacctgggtattgc	78							
QY	252	ccccggggcccccaacctctgtctaa	277							
Db	79	tgcctctcttcacactgcagtgctca	104							

RESULT 12
AAZ45827/C
ID AAZ45827 standard; DNA; 2088 BP.
XX
XX AAZ45827;
XX
XX
XX 25-APR-2000 (first entry)
XX
XX Degenerate DNA encoding a disintegrin homologue designated zdintl.
XX
XX Human; disintegrin homologue; zdintl; cardiac myocyte; adipocyte;
KW gene therapy; cell-cell interaction; chondro sarcoma; atherosclerosis;
KW Alzheimer's disease; restenosis; ischemic reperfusion; obesity;
KW intimal hyperplasia; tumour; platelet aggregation; apoptosis;
KW neurogenesis; myogenesis; arthritis; myogenesis; neurogenesis;
KW connective tissue disorder; chondrogenesis; tumour proliferation;
KW inflammation; ss.
XX
XX Synthetic.
OS
OS Homo sapiens.
XX
XX WO200002912-A2.
XX
XX 20-JAN-2000.
XX
XX 09-JUL-1999; 99WO-US15638.
XX
XX 10-JUL-1998; 98US-0113883.
XX
XX (ZYMO) ZYMOGENETICS INC.
XX
XX Sheppard PO, Baindur N, Deisher TA, Bishop PD;
PI WPI; 2000-160898/14.
XX
XX Polypeptide useful in modulating cell-cell interaction in tissues of
PT heart, brain, spinal cord and treating chondro sarcoma,
PT atherosclerosis, restenosis, obesity, intimal hyperplasia and tumors -
XX Disclosure; Page 129-130; 132pp; English.
PS
XX The present sequence represents a degenerate sequence which encodes
CC a human disintegrin homologue, designated zdintl. The zdintl polypeptide
CC is a cardiac myocyte proliferation and differentiation stimulator, as
CC well as an adipocyte proliferation and differentiation inhibitor.
CC Polynucleotides encoding zdintl are used in gene therapy. The zdintl
CC polypeptide is useful in modulating cell-cell interactions of cells
CC derived from tissues of heart, brain, spinal cord and skeletal muscle.
CC It is useful in treating and diagnosing chondro sarcoma, atherosclerosis,
CC Alzheimer's disease, restenosis, ischemic reperfusion, obesity, intimal
CC hyperplasia and tumors of heart, brain, and spinal cord. The zdintl
CC polypeptide is also useful in identifying its new family members,
CC antagonists, agonists and antibodies. Antagonists, antibodies and fusion
CC proteins of zdintl are useful in inhibiting platelet aggregation,
CC apoptosis, neurogenesis and myogenesis. Agonists and antagonists are
CC useful in studying cell-cell interactions, arthritis, myogenesis,
CC neurogenesis, connective tissue disorders, chondrogenesis, tumour
CC proliferation and suppression, extracellular matrix proteins, repair
CC and remodelling of ischemic reperfusion, inflammation, and apoptosis.
XX
XX Sequence 2088 BP; 412 A; 178 C; 381 G; 262 T; 855 other;

Query Match	10.7%	Score 30	DB 21	Length 2088
Best local Similarity	27.9%	Pred. No. 3.4		
Matches	46	Conservative	36	Mismatches 83
				Indels 0
				Gaps 0
Qy	65	gctgcagtcgacgcgaaccactttccacgtccaccaatccaggtgagtagcttgctggaga	124	
Db	190	GYTTCGCTTTCCTGATGATYTCNACRTARTCSNWSNARNCCRTTTRTTNARDA	131	
Qy	125	agaagcgcgaagcctctatcacgcgtccattgttttgcgcgacgcgcgagctgagtt	184	
Db	130	TNARTCNARDATRAAYTTNSWNCRAANGCYTCDATYTGRAANSWNGCYTGNGCNARRT	71	
Qy	185	gatcggcgcccaaaaacctctgctaagtcgcgctgaagtataaa	229	
Db	70	CNACNGCYTTRTTGTYTGYTGCTCNCKNGCYTTNGTTCNA	26	
RESULT	13			
AAS11992/c				
ID	AAS11992	standard; DNA; 2088 BP.		
XX	AAS11992;			
AC	AAS11992;			
DT	04-DEC-2001	(first entry)		
XX				
DE	Human degenerate DNA for disintegrin protease zdtnt1.			
XX				
KW	Human; disintegrin; zdtnt1; ds; antiangiogenic; vascular; thrombolytic.			
KW	cell matrix; cell-cell interactions; apoptosis; neurogenesis;			
KW	connective tissue disorders; chondrogenesis; arthritis;			
KW	tumour proliferation; ischaemia reperfusion; inflammation;			
KW	chromosome 2q33.			
XX				
OS	Homo sapiens.			
XX				
PN	US6265199-B1.			
XX				
PD	24-JUL-2001.			
XX				
PF	09-JUL-1999;	99US-0351414.		
XX				
PR	10-JUL-1998;	98US-0092371.		
XX				
PA	(ZYMO) ZYMOGENETICS INC.			
XX				
PI	Sheppard PO, Baidur N, Deisher TA, Bishop PD, Taft DW;			
XX				
DR	WPI; 2001-450736/48.			
XX				
PT	Disintegrins protease zdtnt1, useful for producing agents for the			
PT	development of antithrombotic and anti-migration of tumour cells and			
PT	have antiangiogenic activity -			
XX				
PS	Disclosure; Column 57-58; 50pp; English.			
XX				
CC	The invention relates to an isolated novel disintegrin protease family			
CC	member, zdtnt1. Disintegrins bind cell surface molecules, including			
CC	integrins, on the surface of various cells such as platelets,			
CC	fibroblasts, tumour, endothelial, muscle, neuronal, bone and sperm cells.			
CC	Disintegrins are unique and potentially useful tools for investigating			
CC	cell matrix and cell-cell interactions, apoptosis, neurogenesis,			
CC	connective tissue disorders, chondrogenesis, arthritis, tumour			
CC	proliferation, ischaemia reperfusion and inflammation. Additionally, they			
CC	are useful in the development of antithrombotic and anti-migration of			
CC	tumour cells and have antiangiogenic activity. The present sequence			
CC	is a degenerate DNA sequence representing all DNAs that encode human			
CC	disintegrin, zdtnt1, the gene for which is located on			
CC	chromosome 2q33.			
XX				
SO	Sequence 2088 BP; 412 A; 178 C; 381 G; 262 T; 855 other;			

	Query Match	10.7%;	Score 30;	DB 17;	Length 2272;	
	Best Local Similarity	59.3%;	Pred. No. 3.5;			
	Matches 51;	Conservative	0;	Mismatches	35;	Indels 0; Gaps 0;
Qy	192	cgcacaaaacctctgctaagtcccggtgctaagataaatagccggcgggtagtatttac	251			
Db	940	cgcaatgaactctgcataaggcgagtcgccactacacatatagcaggcatgcctggattgc	999			
Qy	252	cccgcgggctccccaaacctctgttaa	277			
Db	1000	tgcctcctttcacactgcagtgctca	1025			

RESULT	15
AAAF74421	
ID	AAAF74421 standard; DNA; 2272 BP.
XX	
XX	
AC	AAAF74421;
XX	
XX	
DT	08-MAY-2001 (first entry)
XX	
XX	
DE	Human TGF-beta MP121 nucleotide sequence SEQ ID NO:3.
XX	
XX	Human; transforming growth factor-beta; monomeric protein; MP52;
KW	MP121; dimeric protein; TGF-beta; vulnery; antiulcer; nootropic;
KW	neuroprotective; antiinfertility; osteopathic; gene therapy; bone;
KW	cartilage; dental; wound healing; connective tissue; ds.

OS				
XX	Homo sapiens.			
XX				
FH	key	Location/Qualifiers		
FH	CDS	138..1186		
FT		/tag=a		
FT		/product= "MP121"		
FT		/note= "transforming growth factor (TGF)-beta monomeric protein"		
FT				
XX				
XX	EP1074620-A1.			
PN				
XX				
XX				
PD	07-FEB-2001.			
XX				
XX				
PF	06-AUG-1999;	99EP-0115613.		
XX				
XX	06-AUG-1999;	99EP-0115613.		
PR				
XX				
XX	(HYGE-) HYGENE AG C/O MAEDER & BAUMGARTNER TREUH.			
PA				

DR WPI; 2001-228100/24.
DR P-PSDB; AAB/0530.
XX
PT Novel monomeric protein of transforming growth factor-beta family for
PT prevention or therapy of diseases associated with bone, cartilage
PT damage, promotion of wound healing, has substitution or deletion of
PT cysteine
XX
XX
PS Claim 16; Page 17-20; 31pp; English.

The present invention describes a protein (I) selected from the members of the transforming growth factor-beta (TGF-beta) superfamily, which is monomeric due to substitution or deletion of a cysteine which is responsible for dimer formation. Also described are: (1) nucleic acid (II) encoding (I); (2) expression vector (III) containing (II) in a suitable vector system; (3) host cell (IV) containing (III) capable of producing (I); and (4) a pharmaceutical composition (V) containing (I), (II), (III) or (IV). (I) has vulnerary, antiulcer, neurotropic, neuroprotective, antifertility and osteopathic activities, and can be used in gene therapy. (V) is useful for the prevention or therapy of diseases for which also the dimeric form of the protein would be indicated. Diseases treatable include diseases associated with bone and/or cartilage damage or affecting bone and/or cartilage disease or situations in which cartilage and/or bone growth is desirable, for spinal

CC fusion, for damaged or diseased tissue associated with connective tissue
 CC including tendon and/or ligament, periodontal or dental tissue including
 CC dental implants, neural tissue including CNS tissue and neuropathological
 CC situations, tissue of the sensory system, liver, pancreas, cardiac,
 CC blood vessel, renal, uterine and thyroid tissue, skin, mucous membrane,
 CC endothelium, epithelium, for promotion or induction of nerve growth,
 CC tissue regeneration, angiogenesis, wound healing including ulcers, burns,
 CC injuries or skin grafts, induction of proliferation of progenitor cells
 CC or bone marrow cells, for maintenance of a state of proliferation or
 CC differentiation, for treatment or preservation of tissue or cells for
 CC organ or tissue transplantation, for integrity of gastrointestinal lining
 CC and for treatment of disturbances in fertility, contraception or
 CC pregnancy. The present sequence encodes the specifically claimed
 CC TGF-beta monomeric protein MP121, from the present invention.
 XX
 SQ Sequence 2272 BP; 510 A; 563 C; 511 G; 585 T; 3 other;

Query Match 10.7%; Score 30; DB 22; Length 2272;
 Best Local Similarity 59.3%; Pred. No. 3.5;
 Matches 51; Conservative 0; Mismatches 35; Indels 0; Gaps 0;
 QY 192 cgccaaacacctgtgtaagtcctgctgaagataaaatagccgggggtagtattac 251
 ||||| ||||| || ||||| || ||||| || ||||| || ||||| ||
 Db 940 cgccatgaacttctgcatagggcagtcgccactacacatagcagcagtcgctgtattgc 999
 QY 252 ccccgcggtccccaacctctgctaa 277
 || | || | || || |
 Db 1000 tgccctcttccacactgcagtgctca 1025

Search completed: June 19, 2002, 17:14:11
 Job time: 18909 sec


```

RESULT      2
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: TUBERCULOSIS
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match          11.5%; Score 32.2; DB 4; Length 4411529;
Best Local Similarity 57.4%; Pred. No. 1.5;
Matches 58; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY   108 gagtgcgttctgagagaacccgaagcctctatctaccgtccattgttttgccg 167
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   3577728 GAGCGCAGCGGGGACAAACACCCCGAGTGCCTGTGCCGTTCGGATCGGTTTGCA 3577669

QY   168 gatcgaggcgtgagttgatctggcgcccaaaacctctgt 208
      | | | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   3577668 GCAC TGTTGCCGAGCGCGGAGGAGCGGAATCCCAACTGCT 3577628

RESULT      3
US-09-462-975-2/c
; Sequence 2, Application US/09462975
; Patent No. 6303345
; GENERAL INFORMATION:
; APPLICANT: Rohde, Wolfgang
; APPLICANT: Becker, Dieter
; APPLICANT: Randies, John W.
; APPLICANT: Hehn, Alain
; APPLICANT: Salamini, Francesco
; TITLE OF INVENTION: USE OF VIRUS DNA AS A PROMOTER
; FILE REFERENCE: 23232.0003U1
; CURRENT APPLICATION NUMBER: US/09/462,975
; CURRENT FILING DATE: 2000-05-17
; PRIOR APPLICATION NUMBER: PCT/EP98/04345
; PRIOR FILING DATE: 1998-07-13
; PRIOR APPLICATION NUMBER: 19730502.4
; PRIOR FILING DATE: 1997-07-16
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 31
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:/note=synthetic
; OTHER INFORMATION: construct
US-09-462-975-2

Query Match          11.0%; Score 31; DB 4; Length 31;
Best Local Similarity 100.0%; Pred. No. 0.0089;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      231 agccgcggggcgtagtattaccgcccgcgct 261

```

```
RESULT 5
US-09-054-526B-5
; Sequence 5, Application US/09054526B
; Patent No. 6197550
; GENERAL INFORMATION:
; APPLICANT: H TTEN, GERTRUD
; APPLICANT: NEIDHARDT, HELGE
; APPLICANT: BECHTOLD, ROLF
; APPLICANT: POHL, JENS
; TITLE OF INVENTION: DNA SEQUENCES ENCODING NOVEL
; TITLE OF INVENTION: GROWTH/DIFFERENTIATION FACTORS
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIKAI DO, MARMEI STEIN, MURRAY & ORAM LLP
; STREET: 655 FIFTEENTH STREET, N. W., G STREET LOBBY,
; STREET: SUITE 330
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-5701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/054.526B
; FILING DATE: 03-APR-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/289,222
; FILING DATE: 12-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE P 44 23 190.3
; FILING DATE: 01-JUL-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EPO 92102324.8
; FILING DATE: 12-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP93/00350
; FILING DATE: 12-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: KITT, MONICA CHIN
; REGISTRATION NUMBER: 36,105
; REFERENCE/DOCKET NUMBER: P564-8005
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202/638-5000
; TELEFAX: 202/638-4810
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 265 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA from mRNA
US-09-054-526B-5

Query Match 10.7%; Score 30; DB 4; Length 265;
Best Local Similarity 59.3%; Pred. No. 0.066;
Matches 51; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

Qy 192 ccccaaaacctgctgaagtcctgctgaagataaaatagccgagggtattac 251
      ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 19 CGCCATGAACCTCTGCTAGGCGAGTCGCCACTACACATAGCAGCATGCTGTGTTTC 78

Qy 252 ccccgaggctcccaacctctgtaa 277
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 79 TGCCTCCTTCACACTGCAGTGCTCA 104

RESULT 6
```

```
US-09-351-414-3/c
; Sequence 3, Application US/09351414
; Patent No. 6265199
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Baindur, Nand
; APPLICANT: Deisher, Theresa A.
; APPLICANT: Bishop, Paul D.
; TITLE OF INVENTION: DISINTEGRIN HOMOLOG
; FILE REFERENCE: 98-29
; CURRENT APPLICATION NUMBER: US/09/351,414
; CURRENT FILING DATE: 1999-07-09
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 2088
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: zdm1 amino acid degenerate sequence
; FEATURE:
; NAME/KEY: variation
; LOCATION: (1)...(2088)
; OTHER INFORMATION: n is any nucleotide
; FEATURE:
; NAME/KEY: misc:feature
; LOCATION: (1)...(2088)
; OTHER INFORMATION: n = A,T,C or G
US-09-351-414-3

Query Match 10.7%; Score 30; DB 4; Length 2088;
Best Local Similarity 27.9%; Pred. No. 0.19;
Matches 46; Conservative 36; Mismatches 83; Indels 0; Gaps 0;

Qy 65 gcttgagtcgacgcaaacacttccacgtccacatccaggtagtagctgtgtgaga 124
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 190 GYTTNCCRTTYTCTARTGDTATTCNACRTARTCNNSNNARNARNCCTRTTNRADA 131

Qy 125 agaaagcgcaagcctctctatctaccgtccattgtttttgctgcgacgagctgatt 184
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 130 TNARCTNARDATRAAYTTNSWNCRAANGCYTCDATYTGRAANSWNGCYTGNGCNARRT 71

Qy 185 gatctggcgcaaaacacctctgtaagtcctgctgaagtataaa 229
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 70 GNACNGCYTTRTYTGYTGYTGYTGYTGYTGYTGYTGYTGYTGYTGYTGYTGYTGYT 26

RESULT 7
US-08-482-577B-1
; Sequence 1, Application US/08482577B
; Patent No. 5807713
; GENERAL INFORMATION:
; APPLICANT: HOTTEN, GERTRUD
; APPLICANT: NEIDHARDT, HELGE
; APPLICANT: BECHTOLD, ROLF
; APPLICANT: POHL, JENS
; TITLE OF INVENTION: DNA SEQUENCES ENCODING NOVEL
; TITLE OF INVENTION: GROWTH/DIFFERENTIATION FACTORS
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIKAI DO, MARMEI STEIN, MURRAY, AND ORAM
; STREET: 655 FIFTEENTH STREET, N.W., G STREET LOBBY,
; STREET: SUITE 330
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
```

PRISON AFFILIATION DATA:

; FILING DATE: 2-FEB-1993
; ESTD: 1992 FEBRUARY 1993

```

; APPLICATION NUMBER: PCT/EP93/00350
; FILING DATE: 12-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: KITTS, MONICA CHIN
; REGISTRATION NUMBER: 36,105
; REFERENCE/DOCKET NUMBER: P564-9021
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202/638-5000
; TELEFAX: 202/638-4810
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2272 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA from mRNA
; US-08-289-222E-2

Query Match 10.7%; Score 30; DB 3; Length 2272;
Best Local Similarity 59.3%; Pred. No. 0.2;
Matches 51; Conservative 0; Mismatches 35; Indels 0;

QY 192 cgcacaaacctctgctaagtcgcgtctaagtaagataataagccgcggggcgtagtat
      ||||| ||||| || ||||| || ||||| || ||||| || |||||
Db 940 CGCATGAATCTTCGATAGGGCAGTGCCTACTACATAGCAGGATGCCTGGTAT

QY 252 ccgcgcgtcccaacctctgctaa 277
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1000 TGCCTCTCTTTACACTGCAGTGCTCA 1025

RESULT 9
US-09-218-176-1
; Sequence 1, Application US/09218176
; Patent No. 6171584
; GENERAL INFORMATION:
; APPLICANT: H TTEN, Gertrud
; APPLICANT: NEIDHARDT, Helge
; APPLICANT: BECHTOLD, Rolf
; APPLICANT: POHL, Jens
; APPLICANT: PAULISTA, Michael
; TITLE OF INVENTION: NEW GROWTH/DIFFERENTIATION FACTORS OF THE
; TITLE OF INVENTION: TGF- FAMILY
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIKARDO, MARCELSTEIN, MURRAY & ORAM LLP
; STREET: 655 Fifteenth Street, N. W., G Street Lobby,
; STREET: Suite 330
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/218,176
; FILING DATE: Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/679,048
; FILING DATE: 12-JUL-1996
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP96/03065
; FILING DATE: 12-JUL-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP93/00350
; FILING DATE: 2-FEB-1993
; PRIOR APPLICATION DATA:

```


RESULT 12

```

US-08-655-836-2/c
; Sequence 2, Application US/08655836
; Patent No. 5824486
; GENERAL INFORMATION:
; APPLICANT: Borden, Laurence A.
; APPLICANT: De Vivo, Michael
; APPLICANT: Yokoyama, Midori
; APPLICANT: Albert, Vivian R.
; TITLE OF INVENTION: GLYCINE TRANSPORTER-TRANSFECTED CELLS
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: P.O. Box 5218
; CITY: Princeton
; STATE: NJ
; COUNTRY: USA
; ZIP: 08543
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patenting Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/655,836
; FILING DATE: 31-MAY-1996
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: Bloom, Allen
; REGISTRATION NUMBER: 29135
; REFERENCE/DOCKET NUMBER: 317743-105
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 609-520-3214
; TELEPHONE: 609-520-3259
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2202 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-655-836-2

```

	Query Match	9.8%	Score 27.4;	DB 1;	Length 2202;
	Best Local Similarity	50.4%;	Pred. No. 2;	Mismatches	Indels
	Matches	67;	Conservative	0;	Gaps
QY	120	ggagaagaaaggccgaacgtctctatcacggtcccattgttttgcgcgatcggcgacct	179		
Dd	269	GAAGAGGGGCGACAGCTGCATGAGCCGGGCCCATCTGTGGCGCAGCATGCGACCCC	210		
QY	180	gagttagtcgtgcgccaaaaccctctgctaagtcctcgtgctaagtataaatagccgcggg	239		
Dd	209	GGTGTCFCGCGGCTCACTTGCACGCTCACAGCGGACACTGACGCATCCCTTCG	150		
QY	240	ggctagtaattacc	252		
Dd	149	CTCTCCCACTGCC	137		

RESULT 13

```

US-09-020-753-2/c
; Sequence 2, Application US/09020753
; Patent No. 5968823
; GENERAL INFORMATION:
; APPLICANT: Borden, Laurence A.
; APPLICANT: Devivo, Michael
; APPLICANT: Yokoyama, Midori
; APPLICANT: Albert, Vivian R.
; TITLE OF INVENTION: Glycine Transporter-Transfected

```

```

; TITLE OF INVENTION: Cells and Uses Thereof
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 997 Lenox Drive, Building 3, Suite 210
; CITY: Lawrenceville
; STATE: NJ
; COUNTRY: USA
; ZIP: 08543
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/020,753
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/655,836
; FILING DATE: 31-MAY-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Bloom, Allen
; REGISTRATION NUMBER: 29,135
; REFERENCE/DOCKET NUMBER: 317743-105
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 609-520-3214
; TELEFAX: 609-520-3259
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2202 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
US-09-020-753-2

Query Match          9.8%; Score 27.4; DB 2; Length 2202;
Best Local Similarity 50.4%; Pred. No. 2;
Matches 67; Conservative 0; Mismatches 66; Indels 0; Gaps

Qy 120 ggagaagaaagccgcgaagcctctatctaocgtccattgttttgcgcatcgacggct 179
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 269 GGAAAGGGGGCCACAGGTCATGAGCGCGGCCCATCTCTGGGCGAGCGATCGCAGCCCC 210
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 180 gacttgatctggcccacaaaaccttcctaagtccccgtctaagtataaatagccgcgg 239
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 209 CGGTGTCCTCCGCCTCATCACCTCTGCCAGTCTCCAGCGGACATGCGCATCCCCTGC 150
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 240 ggcgtagtattacc 252
    || | | | |
Db 149 CTCCTCCCACTGCC 137
    || | | | |

RESULT 14
US-08-655-836-3/c
; Sequence 3, Application US/08655836
; Patent No. 5824486
; GENERAL INFORMATION:
; APPLICANT: Borden, Laurence A.
; APPLICANT: De Vivo, Michael
; APPLICANT: Yokoyama, Midori
; APPLICANT: Albert, Vivian R.
; TITLE OF INVENTION: GLYCINE TRANSPORTER-TRANSFECTED CELLS
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: P.O. Box 5218
; CITY: Princeton
; STATE: NJ
; COUNTRY: USA

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2002, 14:04:28 ; Search time 7489.97 Seconds
(without alignments)
506.363 Million cell updates/sec

Title: US-09-462-955b-1_COPY_711_991

Perfect score: 281

Sequence: 1 aatcgagtattggcgccgccc.....ccccaaacctgtgtaacccc 281

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vit:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	34.4	12.2	898	9	AL667528
C 2	34	12.1	1101	12	CNS00DBS
C 3	33.2	11.8	1030	9	BB610652
C 4	33	11.7	644	12	AZ339575
5	33	11.7	762	9	AV763886
6	33	11.7	977	12	A0459250
C 7	32.8	11.7	668	12	BH437599
8	32.6	11.6	659	9	AA221432
9	32	11.4	1037	12	CNS0374V
10	32	11.4	1210	11	AK010602
11	31.8	11.3	533	9	AW450316
C 12	31.8	11.3	578	10	C79787
C 13	31.6	11.2	335	10	BE246832
14	31.6	11.2	358	9	AI205836
15	31.6	11.2	410	9	AA834925
16	31.4	11.2	416	9	AI016082
17	31.4	11.2	424	9	AA889602

C 18	31.2	11.1	297	9	BB295841
C 19	31.2	11.1	299	9	BB285126
20	31.2	11.1	391	9	AW037599
21	31.2	11.1	617	12	BH087127
22	31	11.0	298	9	AW014292
23	31	11.0	620	12	BH540957
24	31	11.0	625	12	AQ874985
C 25	31	11.0	692	12	AG145182
C 26	31	11.0	694	12	BH451673
27	31	11.0	768	9	AL666434
28	30.8	11.0	440	9	AA438686
29	30.8	11.0	533	10	BJ198148
30	30.8	11.0	713	10	BT552821
31	30.8	11.0	875	10	BE962362
C 32	30.8	11.0	918	10	BE378992
33	30.8	11.0	1101	12	CNS001C3
34	30.6	10.9	228	9	AV308279
C 35	30.6	10.9	319	9	AV104431
36	30.4	10.8	247	12	AQ025371
37	30.4	10.8	260	9	A1336966
38	30.4	10.8	260	9	A1337129
39	30.4	10.8	260	9	A1624907
40	30.4	10.8	289	9	AI825120
C 41	30.4	10.8	340	9	AA051125
C 42	30.4	10.8	373	12	AZ812694
C 43	30.4	10.8	389	12	AZ510706
C 44	30.4	10.8	633	9	AV689499
45	30.4	10.8	795	12	AQ876011

ALIGNMENTS

RESULT 1

AL667528
LOCUS AL667528 898 bp mRNA linear EST 11-JAN-2002
DEFINITION clone 020ZG05 5', mRNA sequence.
ACCESSION AL667528.1 GI:18134435
VERSION AL667528
KEYWORDS EST.
SOURCE Ciona intestinalis.
ORGANISM Ciona intestinalis.
Fukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona; Phlebobranchia; Cionidae; Ciona.
REFERENCE 1 (bases 1 to 898)
AUTHORS Genoscope.
TITLE Ciona intestinalis directional larval cDNA library
JOURNAL Unpublished (2002)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequençage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
IMPORTANT: this sequence may contain errors. The Ciona intestinalis library from which the clone was isolated may be contaminated with cDNAs from bacteria or other Eukarya.
Directional larval cDNA library originate from Dr M.Branno, Stazione A.Dohrn, Naples, Italy, and was prepared in pBluescript2SK+.

FEATURES

source
1. 898
/organism="Ciona intestinalis"
/db_xref="taxon:7719"
/clone="020ZG05"
/clone.lib="directional larval cDNA library"
/note="Vector: pBluescript2SK+" 1 others
289 a 186 c 157 g 265 t

BASE COUNT

ORIGIN
Query Match 12.2%; Score 34.4; DB 9; Length 898;
Best Local Similarity 60.9%; Pred. No. 1.5;
Matches 56; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

<p>electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gil14732114 gb AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent <i>E. coli</i> XL10-Gold (Stratagene) cells and selected for ampicillin resistance.</p>									
BASE COUNT	177 a	186 c	100 g	181 t					
ORIGIN									
Query Match	11.7%	Score 33;	DB 12;	Length 644;					
Best Local Similarity	60.7%	Pred. No. 4.1;							
Matches	54;	Conservative	0;	Mismatches 35;	Indels	0;	Gaps	0;	
Qy	58	ttactttgcttcagtcacgcacaccatttcacgctcaccatccaggtgagtagcttg	117						
Db	354	TTTCTGTGCACTAGTGTAGTGTATCTCTGCTCCCTTACAAGCCTAAGGCATTCTCTGG	295						
Qy	118	ctggagagaagaagcgcgaagcctctatct	146						
Db	294	CTTTAATAAGAGCTGCAGCCATATATCT	266						
RESULT	5								
AV763886									
LOCUS	AV763886	762 bp	mRNA	linear	EST 19-OCT-2000				
DEFINITION	AV763886	MDS Homo sapiens cDNA clone MDSEEE03 5', mRNA sequence.							
ACCESSION	AV763886								
VERSION	AV763886.1	GI:10921734							
KEYWORDS	EST.								
SOURCE	human.								
ORGANISM	Homo sapiens								
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;								
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.								
	1 (bases 1 to 762)								
	Gu, J., Zhao, M., Huang, Q., Xu, X., Li, Y., Peng, Y., Song, H., Xiao, H.,								
	Gu, Y., Li, N., Qian, B., Liu, F., Qu, J., Gao, X., Cheng, Z., Xu, Z., Zeng								
	, L., Xu, S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G.,								
	Yang, Y., Gao, G., Zhang, Q., Chen, S., Han, Z. and Chen, Z.								
TITLE	Homo sapiens cDNA MDS clones								
JOURNAL	Unpublished (2000)								
COMMENT	Contact: zeguang Han								
	Chinese National Human Genome Center at Shanghai								
	351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai								
	201203, P. R. China								
	Tel: 86-21-50801919(ex.45)								
	Fax: 86-21-50801922								
	Email: hanzg@chgc.sh.cn								
	This clone is available at CHGC in Shanghai.								
FEATURES	Location/Qualifiers								
source	1..762								
	/organism="Homo sapiens"								
	/db_xref="taxon:9606"								
	/clone="MDSEEE03"								
	/clone.lib="MDS"								
	/tissue_type="Bone marrow"								
	/cell_type="CD34+ hematopoietic stem/progenitor cell"								
	/lab_host="BM25.8"								
	/note="Vector: pTriplex2; Site_1: sfIIA; Site_2: sfIIB"								
BASE COUNT	203 a	172 c	177 g	209 t	1 others				
ORIGIN									
Query Match	11.7%	Score 33;	DB 9;	Length 762;					
Best Local Similarity	48.9%	Pred. No. 4.3;							
Matches	87;	Conservative	0;	Mismatches 91;	Indels	0;	Gaps	0;	
Qy	36	catttgatagaagacgaatctgttactttgcttcagtcagcgaacacatttcacgctc	95						
Db	471	CAGCTTGTCGNGAAGTTACCTGATTGTTATTATTAGAATGATACCACTCTGCTGATTC	530						


```

); NCI_CGAP_GC4 pool 1 LLAM 3164-3167, 3716-3720,
3733-3735 (IMAGE Clonoids 1257096-1258631,1469064-1470983,
1475592-1476743); NCI_CGAP_Pr22 pool 1 LLAM 2457-2459,
2758-2759, 3062-3068 (IMAGE Clonoids 985608-986759
1101192-1101959, 1217928-1220615); NCI_CGAP_Co10 pool 1
LLAM 2644-2653, 2871-2872 (IMAGE Clonoids 1057416-1061255
,1144584-1145351). (10% of the driver population), plus a
pool of 3,840 arrayed clones from NCI_CGAP_Sub1 (IMAGE
Clonoids 2708616-2710535) and NCI_CGAP_Sub2 (IMAGE
Clonoids 2710536-2712455) (10% of the driver population
), plus a pool of 11,136 clones from NCI_CGAP_Sub3 (IMAGE
Clonoids 2712456-2723591) (10% of the driver population),
plus a pool of 5,472 clones from NCI_CGAP_Sub4 (IMAGE
Clonoids 2723592-2728969) (70% of the driver population).
Subtraction was performed as previously described [Bonaldi
, Lennon & Soares (1996): Normalization and Subtraction:
Two Approaches To Facilitate Gene Discovery. Genome
Research 6, 791-806.
TAG_LIB=NCI_CGAP_GC4
TAG_TISSUE=germ cell
TAG_SEQ=AAATC"
BASE COUNT      95 a 144 c 146 g 147 t 1 others
ORIGIN

```

```

Query Match      11.3%; Score 31.8; DB 9; Length 533;
Best Local Similarity 53.7%; Pred. No. 9.7;
Matches 66; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

QY 33 ttccatttggaagaacgaacttcttacttctgttcagtcgacgaaccactttccac 92
|| ||||| ||||| || ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 10 TTTTTCGGGAGAAATGTGTCATTTCTGAATCCCTTCTCCACTGACCCACATTCAT 69

QY 93 gtccaccatccaggtgagtagcttctgtgagaagaagccgcagcctctctaccgtc 152
|| || || || || || || || || || || || || || || || || || || || ||
Db 70 CCCTCCACACACTTCAGTAGTGTGCTTACTGCTGCTCAGGCCCTTGACCCCGAGC 129

QY 153 cat 155
Db 130 TAT 132

```

```

RESULT 12
C79787/c
LOCUS
DEFINITION
C79787 Mouse 3.5-dpc blastocyst cdna Mus musculus cDNA clone
J0071H05 3', mRNA sequence.
C79787
EST.
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 578)
Ko, M.S.H., Kitchen, J.R., Wang, X., Wang, X., Threat, T.A., Sun, T.,
Grahovac, M.J., Mason, S., Lim, M.K., Paonessa, P.D., Sauls, A.D. and
Doi, H.
Systematic analyses of genes expressed in 3.5-dpc mouse blastocyst
(The ERATO/Doi Project at Wayne State University)
Unpublished (1997)
Contact: Hirofumi Doi
Doi Biosymetry Project, ERATO
Japan Science and Technology Corporation (JST)
WBG Marine East 12F, 2-6 Nakase, Mihama-ku, Chiba 261-71, Japan
Email: hdoibio.jst.go.jp.
Location/Qualifiers
1 . 578
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="J0071H05"
/clone_lib="Mouse 3.5-dpc blastocyst cdna"

```

```

/tissue_type="blastocyst"
/dev_stage="3.5-dpc"
BASE COUNT      171 a 134 c 113 g 158 t 2 others
ORIGIN
Query Match      11.3%; Score 31.8; DB 10; Length 578;
Matches 63; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 65 gcttgagtcgacgcaaccactttccacgtccacatccaggtgagtagcttctctgaga 124
|| || || || || || || || || || || || || || || || || || || || ||
Db 430 GCTGGCCTTAGCACACACACCTTTTCCCTCTACTAGAGCTTGGCTTGAGGAAG 371

QY 125 agaaagccgcaagccctctatctaccgtccattgttttgcgcgacgcagcgctg 180
|| || || || || || || || || || || || || || || || || || || || ||
Db 370 ATGAAGAGCAGCGCTTCTGTTTACTAGAAAAATGTTTGGTTTCACAGCGTGCTG 315

RESULT 13
BE246832/c
LOCUS
DEFINITION
TCBAP15158 Pediatric pre-B cell acute lymphoblastic leukemia
Baylor-HGSC project=TCBA Homo sapiens cDNA clone TCBAP5158, mRNA
sequence.
ACCESSION
BE246832
VERSION
BE246832.1 GI:9098582
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 335)
Wei, X., Tsang, Y.T.M., Mei, G., Ku, J.M., Ali-Osman Jr., F.R., Muzny, D.,
Bouck, J., Gibbs, R.A. and Margolin, J.F.
Pediatric Leukemia cDNA Sequencing Project
Unpublished (2000)
Contact: Dr. Judith F. Margolin
Texas Children's Cancer Center and Human Genome Sequencing Center
at Baylor College of Medicine
1102 Bates, MC3-3320 Houston, TX 77030, USA
Tel: 832-824-4536
Fax: 832-825-4038
Email: clones@ccc.org
Citation: Carninci, P. and Hayashizaki, Y. High efficiency
full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
Seq primer: M13 primer.
Location/Qualifiers
1 . 335
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="TCBAP5158"
/clone_lib="Pediatric pre-B cell acute lymphoblastic
leukemia Baylor-HGSC project=TCBA"
/sex="male"
/tissue_type="leukopheresis"
/cell_type="pre-B cell"
/dev_stage="pediatric 2 years"
/lab_host="DH10B"
/notes="Vector: lambda pSB; Site_1: BamHI; Site_2: EcoRI;
First strand cDNA was primed with an anchored
XhoI-oligo(dT) primer [5'GGAGACTCGAGCGCGCCGAGGAG(T)VN
3'; V=A,C,G; N=A,C,G,T] and then dG tailed. Second strand
was primed with a BamHI-dC primer
[5'AGAGAGTCGATCCGCGCGCGCAATAATAAT(C) 3'].
Double-stranded cDNA was then digested with BamHI and XhoI
and directionally cloned into the BamHI and SalI sites of
lambda pSB vector. Library went through one round of
normalization. Library was constructed by Wei Yu at RIKEN
of Japan (Carninci, P., Westover, A., Nishiyama, Y., Ohsumi, T.,
Itoh, M., Nagaoka, S., Sasaki, N., Muramatsu, M.,
Schneider, C., Hayashizaki, Y. High efficiency selection of
full-length cDNA by improved biotinylated cap trapper.,

```


Qy 98 caatccaggtgagtagcttgctggagaaagcgcgcaagcctctatctaccgtccattt 157
 Db 61 CACACAACCTTCAGTAGGTTGATGGCTTCTACTGGTCCAGGCCCTTGACCCCGAGCTAGGG 120
 Qy 158 gtttttgcgcgacgcgacggt 179
 Db 121 CCTCAGGGACCTTCGGGCGTT 142

Search completed: June 19, 2002, 14:04:31
 Job time: 7529 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2002, 15:49:33 ; Search time 5904.86 seconds
(without alignments)
184.285 Million cell updates/sec

Title: US-09-462-955B-1_COPY_734_785
Perfect score: 52
Sequence: 1 aataaagattccatttgga.....tggtactttgttgagtcgc 52

Scoring table: IDENTITY NUC
Gapop 10_0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_ov.*

5: gb_pat.*

6: gb_ph.*

7: gb_pl.*

8: gb_ro.*

9: gb_sy.*

10: gb_un.*

11: gb_vl.*

12: gb_vl.*

13: gb_vl.*

14: gb_vl.*

15: gb_vl.*

16: gb_vl.*

17: gb_vl.*

18: gb_vl.*

19: gb_vl.*

20: gb_vl.*

21: gb_vl.*

22: gb_vl.*

23: gb_vl.*

24: gb_vl.*

25: gb_vl.*

26: gb_vl.*

27: gb_vl.*

28: gb_vl.*

29: gb_vl.*

30: gb_vl.*

31: gb_vl.*

32: gb_vl.*

33: gb_vl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No. Score Match Length DB ID Description

ALIGNMENTS

RESULT 1

ARI72770

LOCUS

ARI72770

DEFINITION

Sequence 1 from patent US 6303345.

ARI72770

ACCESSION

ARI72770.1

VERSION

GI:17912261

KEYWORDS

Unknown.

SOURCE

Unknown.

ORGANISM

Unclassified.

REFERENCE

1 (bases 1 to 1291)

AUTHORS

Rohde,W., Becker,D., Randles,J.W., Hehn,A. and Salamini,F.

TITLE

Use of a virus DNA as promoter

JOURNAL

Patent: US 6303345 A 1 16-OCR-2001;

FEATURES

Location/Qualifiers

1..1291

/organism="unknown"

source

BASE COUNT

336 a 323 c 332 g 300 t

ORIGIN

Query Match 100.0%; Score 52; DB 6; Length 1291;

Best Local Similarity 100.0%; Pred. No. 5.6e-07;

ARI72770 Sequence
M29963 Coconut fol
AP004548 Lotus jap
AX036735 Sequence
X52938 Commelina y
Z81038 Caenorhabdi
AL135782 Homo sapi
AL008632 Human DNA
AC105955 Mus muscu
AC139381 Human DNA
AC108771 Homo sapi
AC018764 Homo sapi
AC090825 Homo sapi
AC096094 Rattus no
AC013519 Homo sapi
AC008990 Homo sapi
AC009009 Homo sapi
AC104068 Homo sapi
AL355350 Human DNA
AC010620 Homo sapi
AC018569 Homo sapi
AC078825 Homo sapi
AC103895 Papio cyn
AC104795 Homo sapi
AC097226 Papio cyn
AL161646 Human DNA
AC044882 Homo sapi
AB040540 Clostridi
AB040712 Clostridi
AB040541 Clostridi
AB040713 Clostridi
AB040542 Clostridi
AB040714 Clostridi
L41647 Drosophila
AL670195 Mus muscu
AL391356 Homo sapi
AL358779 Human DNA
AL356114 Human DNA
AL445683 Human DNA
U89127 Sus scrofa
AP004093 Oryza sat
AC055742 Homo sapi
AC073574 Homo sapi
AC019285 Homo sapi
AL353741 Human DNA

Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 aataaagattccattgataagaacgaatctgttaattgcttgcagtc 52
 |||||
 Db 1124 AATAAAGATCCATTGGATAAGAACGAATCTGTTACTTGTTCGTCAGTC 1175

RESULT 2

CFDCG
 LOCUS Coconut foliar decay virus, complete genome.
 DEFINITION Coconut foliar decay virus, complete genome.
 ACCESSION M29963
 VERSION M29963.1 GI:323306

KEYWORDS circular; complete genome.
 SOURCE Coconut foliar decay virus DNA, clones NMORG, PORG, and PORGSUA3A.
 ORGANISM Coconut foliar decay virus

REFERENCE 1 (bases 1 to 1291)
 Viruses; ssDNA viruses; Nanovirus.

AUTHORS Rohde, W., Randles, J.W., Langridge, P. and Hanold, D.

TITLE Nucleotide sequence of a circular single-stranded DNA associated

JOURNAL with coconut foliar decay virus

MEDLINE Virology 176, 648-651 (1990)

COMMENT Draft entry and printed sequence for [1] kindly submitted by

W. Rohde, 15-MAR-1989, for release after publication.

FEATURES

Source

1. 1291

/organism="Coconut foliar decay virus"

/db_xref="taxon:12474"

40. .70

/note="stem-loop structure"

103. .975

/note="ORF 1"

/codon_start=1

/protein_id="AAA42894.1"

/db_xref="GI:323307"

/translation="MGSSIRWCFLLNYETEEAANVRRRIESLNLYAIVGDEVAPS

TRGRLHQFIHLKTRGRRLQGLKTLVGNDRILHLEPRGDEQNRDYSKERVLLLEHGV

TPGVKRAFLAQRFAEDELRLDLEPGYRCVHVHGSVETRWAAENPFPPYHNWQ

LEVLSAIGEPADRLILATCGRDGDKSVAKYLGKLPDMFTYTCGGPRKDVLYQIE

DKRNLILDVPCRNLELYNALLECVKRNRAESSDKYEPILYFGPDHVLVFPANVLPD

YLRNDRIRKLWNI

314. .775

/note="ORF 2"

/codon_start=1

/protein_id="AAA42895.1"

/db_xref="GI:323308"

/translation="MTGFTWSPVVPVTHRIETTVRRNGCFSTESRLVLESKHDPN

DLRLNLSAWKQADTEDALYELRWNGQDGLKIRSHFHTITIGSLKCCLRSESQRT

IAQSGYADETEETSPCLPNISDSPTGSHVPEPTYCTSTRTONEI

complement(422. .568)

/note="ORF 6"

/codon_start=1

/protein_id="AAA42896.1"

/db_xref="GI:323309"

/translation="MEMGTDFQRPILSIPPKLRVORIFGIRLPGGVHVPQQIVGP

IVAF

639. .797

/note="ORF 3"

/codon_start=1

/protein_id="AAA42897.1"

/db_xref="GI:323310"

/translation="MTRRRRRREVRCVQISRTQARVLHMMWNQKGRIVPVHRGPKTK

FNPRCTQV

complement(823. .987)

/note="ORF 5"

/codon_start=1

/protein_id="AAA42898.1"

/db_xref="GI:323311"

/translation="MPTLNIPOFYSVPADFQIIRODIGKEYEYMHMVEPKITKGFVFF

RTECPVLNTF

1098. .1286

/note="ORF 4"

/codon_start=1
 /protein_id="AAA42899.1"
 /db_xref="GI:323312"
 /translation="MNRVMGGPTIKDSIWIRNLLCLQCTQPLSTSPIQVSSLEKK
 AASLYLPSICFCAIGRLS"

BASE COUNT 336 a 323 c 332 g 300 t
 ORIGIN

Query Match 100.0%; Score 52; DB 14; Length 1291;
 Best Local Similarity 100.0%; Pred. No. 5,6e-07;
 Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 aataaagattccattgataagaacgaatctgttaattgcttgcagtc 52
 |||||
 Db 1124 AATAAAGATCCATTGGATAAGAACGAATCTGTTACTTGTTCGTCAGTC 1175

RESULT 3

AP004548/c

LOCUS

DEFINITION

Lotus japonicus genomic DNA, chromosome 6, clone: LjT31L24, TM0228b,

complete sequence.

ACCESSION AP004548

VERSION AP004548.1 GI:17736915

KEYWORDS HTG.

SOURCE

ORGANISM

Lotus japonicus DNA, clone_lib:LjT library clone:LjT31L24.

Lotus japonicus

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Lotaeae;

Lotus.

REFERENCE 1 (sites)

AUTHORS Sato, S., Kaneko, T., Nakamura, Y., Asamizu, E., Kato, T. and Tabata, S.

TITLE Structural Analysis of a Lotus japonicus Genome. I. Sequence

Features and Mapping of fifty-six TAC clones which cover the 5.4 Mb

Regions of the Genome

Unpublished

REFERENCE 2 (bases 1 to 19059)

AUTHORS Nakamura, Y.

TITLE Direct Submission

JOURNAL Submitted (13-DEC-2001) Yasukazu Nakamura, Kazusa DNA Research

Institute, Department of Plant Gene Research; 1532-3, Yana,

Kisarazu, Chiba 292-0812, Japan (E-mail: ynakamuekazusa.or.jp,

URL: http://www.kazusa.or.jp, Tel: 81-438-52-3935,

Fax: 81-438-52-3934)

FEATURES

Location/Qualifiers

1. 19059

/organism="Lotus japonicus"

/db_xref="taxon:34305"

/chromosome="6"

/clone="LjT31L24"

/clone_lib="LjT library"

/note="TM0228b, a part of TAC clone: TM0228"

BASE COUNT 6104 a 3232 c 3509 g 6214 t

ORIGIN

Query Match 54.2%; Score 28.2; DB 8; Length 19059;

Best Local Similarity 80.5%; Pred. No. 21;

Matches 33; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 aataaagattccattgataagaacgaatctgttaattgcttgcagtc 41

|||||

Db 18358 AATAAAGATTACATTGATAAGATCCACTCTGTTTTTT 18318

RESULT 4

AX036735/c

LOCUS

DEFINITION

Sequence 1 from Patent WO0058485.

ACCESSION AX036735

VERSION AX036735.1 GI:11226244

AX036735 243 bp DNA linear PAT 16-NOV-2000

[illegible]

* by the finished sequence as soon as it is available and
 * the accession number will be preserved.

FEATURES

source

```
1..63682
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/map="q21.1-21.33"
/clone="RP3-346A12"
/clone_lib="RPCI-3"
misc_feature
1..63682
/note="assembly_fragment:00926"
BASE COUNT 20610 a 11146 c 11344 g 20582 t
ORIGIN
```

Query Match 50.8%; Score 26.4; DB 2; Length 63682;
 Best Local Similarity 75.0%; Pred. No. 72;
 Matches 33; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 4 aaagattccattgataagaacgaatctgttactttgtctgc 47
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 58003 AAAATAATACATTGGTAGGTGCTAATCTGTACTGTATTGC 58046

RESULT 8

HS321115

LOCUS

Human DNA sequence from PAC 321115 on chromosome Xq21.3.

ACCESSION

AL008632

VERSION

AL008632.1

KEYWORDS

Xq21.3.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

Mistry,S.

Direct Submission

Submitted (28-OCT-1997) Chromosome X Project Group

(http://www.sanger.ac.uk/HGP/ChrX/) Sanger Centre, Hinxton,

Cambridgeshire, CB10 1SA, UK. E-mail enquires:

humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

On Apr 8, 1998 this sequence version replaced gi:2578085.

IMPORTANT: This sequence is the entire insert of clone 321115.

During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the

variations annotated may not be found in the sequence submission

corresponding to the overlapping clone as we submit sequences with

only a small overlap as described above.

This sequence was generated from part of bacterial clone contigs of

human chromosome X, constructed by the Sanger Centre chromosome X

mapping group. Further information can be found at

http://www.sanger.ac.uk/HGP/ChrX/

This sequence has been finished according to sequence map criteria

as follows. An attempt is made to resolve all sequencing problems,

such as compressions and repeats, but not necessarily within known

annotated human repeat sequence elements (e.g. Alu). Where the

sequence is ambiguous, there is an annotation using the 'unsure'

feature key.

The true left end of clone 321115 is at 1 in this sequence. The

true right end of clone 321115 is at 75793.

321115 is from the library RPC11 constructed at the Roswell Park

Cancer Institute by the group of Pieter de Jong.

For further details see http://bacpac.med.buffalo.edu/.

Location/Qualifiers

1..75793

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="X"

/map="Xq21.3"

/clone="RP1-321115"

FEATURES

source

```
repeat_region
1184..1696
/note="LIPAS repeat: matches 378..892 of consensus"
2499..2800
/note="ALUSq repeat: matches 1..302 of consensus"
4436..4469
/note="17 copies of 2 mer 82 % conserved"
5196..5346
/note="MER4A2 repeat: matches 1..152 of consensus"
5347..5646
/note="MER4A2 repeat: matches 211..506 of consensus"
6135..6220
/note="MER34 repeat: matches 530..450 of consensus"
7224..7954
/note="LTR8 repeat: matches 690..1 of consensus"
9239..9272
/note="17 copies of 2 mer 82 % conserved"
9487..9788
/note="ALUSq repeat: matches 1..302 of consensus"
10896..11248
/note="THE1C repeat: matches 370..1 of consensus"
11987..12612
/note="LTR8 repeat: matches 1..558 of consensus"
13703..14143
/note="MER4B repeat: matches 611..194 of consensus"
14353..14597
/note="MER4B repeat: matches 219..1 of consensus"
16357..16402
/note="23 copies of 2 mer 89 % conserved"
16825..16950
/note="ALUJo repeat: matches 136..11 of consensus;
incomplete repeat"
18670..18972
/note="ALUJo repeat: matches 301..2 of consensus"
19098..19184
/note="ALUJo repeat: matches 1..87 of consensus;
incomplete repeat"
19189..19340
/note="MER39 repeat: matches 311..460 of consensus"
19990..20292
/note="ALUSq repeat: matches 1..303 of consensus"
21851..21989
/note="LIPB2 repeat: matches 183..41 of consensus"
21851..21953
/note="LIMD2 repeat: matches 183..81 of consensus"
22141..22409
/note="LIME3 repeat: matches 911..635 of consensus"
23025..23532
/note="2 copies of 254 mer 100 % conserved"
23884..24289
/note="MTLD repeat: matches 57..503 of consensus"
26286..26319
/note="17 copies of 2 mer 88 % conserved"
26910..26955
/note="23 copies of 2 mer 83 % conserved"
28242..28531
/note="ALUy repeat: matches 1..290 of consensus"
29316..29645
/note="ALUJo repeat: matches 302..1 of consensus"
30847..30979
/note="FLAM_C repeat: matches 133..1 of consensus"
30997..31030
/note="17 copies of 2 mer 82 % conserved"
32102..32229
/note="ALUJo repeat: matches 129..1 of consensus;
incomplete repeat"
32751..33112
/note="MT1A2 repeat: matches 1..374 of consensus"
33120..33192
/note="MER5B repeat: matches 155..84 of consensus"
33812..34808
/note="L1 repeat: matches 4372..5384 of consensus"
34668..35665
```

```

repeat_region /note="L1MA4 repeat: matches 4. .1046 of consensus"
36335. .36634
repeat_region /note="AluJo repeat: matches 1. .302 of consensus"
38132. .38448
repeat_region /note="MER48 repeat: matches 611. .297 of consensus"
38453. .38682
repeat_region /note="MER48 repeat: matches 231. .1 of consensus"
39063. .39247
repeat_region /note="Aluub repeat: matches 1. .177 of consensus;
incomplete repeat"
39702. .39837
repeat_region /note="Aluub repeat: matches 166. .301 of consensus;
incomplete repeat"
39884. .40234
repeat_region /note="L1T1A1 repeat: matches 365. .1 of consensus"
40875. .41053
repeat_region /note="MER5B repeat: matches 1. .173 of consensus"
41288. .41418
repeat_region /note="AluJo repeat: matches 132. .1 of consensus;
incomplete repeat"
41628. .41914
repeat_region /note="L1PB3 repeat: matches 896. .591 of consensus"
41746. .42317
repeat_region /note="L1MA5 repeat: matches 758. .136 of consensus"
42058. .42427
repeat_region /note="L1PB2 repeat: matches 395. .29 of consensus"
42411. .44898
repeat_region /note="L1 repeat: matches 5377. .2924 of consensus"
44915. .45196
repeat_region /note="AluJo repeat: matches 17. .293 of consensus;
incomplete repeat"
45240. .45848
repeat_region /note="L1 repeat: matches 2875. .2257 of consensus"
45872. .46170
repeat_region /note="Aluub repeat: matches 5. .302 of consensus"
46258. .46557
repeat_region /note="AluJo repeat: matches 300. .1 of consensus"
46572. .47422
repeat_region /note="L1 repeat: matches 2198. .1356 of consensus"
47151. .47767
repeat_region /note="MER25 repeat: matches 2135. .1510 of consensus"
48136. .48214
repeat_region /note="MSTA repeat: matches 79. .1 of consensus"
48215. .49832
repeat_region /note="MST-INTERNAL repeat: matches 1651. .1 of consensus"
49833. .50232
repeat_region /note="MER25 repeat: matches 1504. .1297 of consensus"
50697. .50903
repeat_region /note="MER25 repeat: matches 1296. .1079 of consensus"
51475. .51694
repeat_region /note="AluX repeat: matches 297. .1 of consensus"
51723. .52016
repeat_region /note="MER25 repeat: matches 1057. .602 of consensus"
52028. .52484
repeat_region /note="MER25 repeat: matches 548. .49 of consensus"
52672. .53166
repeat_region /note="MER25 repeat: matches 128. .11 of consensus"
53615. .53733
repeat_region /note="FLAM_A repeat: matches 128. .11 of consensus"
53875. .54925
repeat_region /note="THE1B-INTERNAL repeat: matches 107. .1167 of
consensus"
54929. .55248
repeat_region /note="THE1C repeat: matches 54. .368 of consensus"
55252. .56389
repeat_region /note="L1 repeat: matches 3226. .2078 of consensus"
56380. .56619
repeat_region /note="L1MB5 repeat: matches 242. .2 of consensus"
56472. .57179
repeat_region /note="L1 repeat: matches 5390. .4668 of consensus"
57183. .57486
repeat_region /note="AluSq repeat: matches 1. .303 of consensus"
57509. .59386

```

```

repeat_region /note="L1 repeat: matches 4680. .2773 of consensus"
59391. .59423
repeat_region /note="L1 copies of 3 mer 85 % conserved"
59435. .59746
repeat_region /note="AluJo repeat: matches 1. .302 of consensus"
59770. .59802
repeat_region /note="AluX/g repeat: matches 36. .4 of consensus;
incomplete repeat"

Query Match 50.8%; Score 26.4; DB 9; Length 75793;
Best Local Similarity 75.0%; Pred. No. 71;
Matches 33; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 4 aaagattccatttggaataagcaagatctgttactttgttc 47
||||| ||||||| ||| ||||||||||| |||||
Db 27259 AAATAATACATTGGTAGGTGCTAATCTGTTACTGTATTGC 27302

RESULT 9
AC105955 LOCUS AC105955 65370 bp DNA linear HTG 11-JAN-2002
DEFINITION Mus musculus clone RP24-262D23, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC105955
VERSION AC105955.1 GI:18129455
KEYWORDS HTG: HTGS, PHASE0.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 65370)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Mus musculus, clone RP24-262D23
Unpublished
2 (bases 1 to 65370)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepl,Y., Collangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
Glade,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Katat,A., Karatas,A., Kells,C., LaRoque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meidrim,J.,
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuppach,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (11-JAN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Snit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www.seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L19374
Center clone name: 262_D_23
-----
* NOTE: This record contains 82 individual
* sequencing reads that have not been assembled into

```

* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

* 1 687: contig of 687 bp in length
* 688 787: gap of 100 bp
* 788 1478: contig of 691 bp in length
* 1479 1578: gap of 100 bp
* 1579 2270: contig of 692 bp in length
* 2271 2370: gap of 100 bp
* 2371 3090: contig of 720 bp in length
* 3091 3190: gap of 100 bp
* 3191 3884: contig of 694 bp in length
* 3885 3984: gap of 100 bp
* 3985 4681: contig of 697 bp in length
* 4682 4781: gap of 100 bp
* 4782 5481: contig of 700 bp in length
* 5482 5581: gap of 100 bp
* 5582 6290: contig of 709 bp in length
* 6291 6390: gap of 100 bp
* 6391 7051: contig of 661 bp in length
* 7052 7151: gap of 100 bp
* 7152 7845: contig of 694 bp in length
* 7846 7945: gap of 100 bp
* 7946 8642: contig of 697 bp in length
* 8643 8742: gap of 100 bp
* 8743 9427: contig of 685 bp in length
* 9428 9527: gap of 100 bp
* 9528 10213: contig of 686 bp in length
* 10214 10313: gap of 100 bp
* 10314 11004: contig of 691 bp in length
* 11005 11104: gap of 100 bp
* 11105 11813: contig of 709 bp in length
* 11814 11913: gap of 100 bp
* 11914 12615: contig of 702 bp in length
* 12616 12715: gap of 100 bp
* 12716 13443: contig of 728 bp in length
* 13444 13543: gap of 100 bp
* 13544 14236: contig of 693 bp in length
* 14237 14336: gap of 100 bp
* 14337 15034: contig of 698 bp in length
* 15035 15134: gap of 100 bp
* 15135 15823: contig of 689 bp in length
* 15824 15923: gap of 100 bp
* 15924 16603: contig of 680 bp in length
* 16604 16703: gap of 100 bp
* 16704 17403: contig of 700 bp in length
* 17404 17503: gap of 100 bp
* 17504 18190: contig of 687 bp in length
* 18191 18290: gap of 100 bp
* 18291 18991: contig of 701 bp in length
* 18992 19091: gap of 100 bp
* 19092 19786: contig of 695 bp in length
* 19787 19886: gap of 100 bp
* 19887 20596: contig of 710 bp in length
* 20597 20696: gap of 100 bp
* 20697 21411: contig of 715 bp in length
* 21412 21511: gap of 100 bp
* 21512 22228: contig of 717 bp in length
* 22229 22328: gap of 100 bp
* 22329 23040: contig of 712 bp in length
* 23041 23140: gap of 100 bp
* 23141 23835: contig of 695 bp in length
* 23836 23935: gap of 100 bp
* 23936 24617: contig of 682 bp in length
* 24618 24717: gap of 100 bp
* 24718 25407: contig of 690 bp in length
* 25408 25507: gap of 100 bp

* 25508 26160: contig of 653 bp in length
* 26161 26260: gap of 100 bp
* 26261 26984: contig of 724 bp in length
* 26985 27084: gap of 100 bp
* 27085 27802: contig of 718 bp in length
* 27803 27902: gap of 100 bp
* 27903 28619: contig of 717 bp in length
* 28620 28719: gap of 100 bp
* 28720 29439: contig of 720 bp in length
* 29440 29539: gap of 100 bp
* 29540 30226: contig of 687 bp in length
* 30227 30326: gap of 100 bp
* 30327 31023: contig of 697 bp in length
* 31024 31123: gap of 100 bp
* 31124 31824: contig of 701 bp in length
* 31825 31924: gap of 100 bp
* 31925 32600: contig of 676 bp in length
* 32601 32700: gap of 100 bp
* 32701 33385: contig of 685 bp in length
* 33386 33485: gap of 100 bp
* 33486 34173: contig of 688 bp in length
* 34174 34273: gap of 100 bp
* 34274 34967: contig of 694 bp in length
* 34968 35067: gap of 100 bp
* 35068 35762: contig of 695 bp in length
* 35763 35862: gap of 100 bp
* 35863 36565: contig of 703 bp in length
* 36566 36665: gap of 100 bp
* 36666 37383: contig of 718 bp in length
* 37384 37483: gap of 100 bp
* 37484 38207: contig of 724 bp in length
* 38208 38307: gap of 100 bp
* 38308 39025: contig of 718 bp in length
* 39026 39125: gap of 100 bp
* 39126 39824: contig of 699 bp in length
* 39825 39924: gap of 100 bp
* 39925 40634: contig of 710 bp in length
* 40635 40734: gap of 100 bp
* 40735 41430: contig of 696 bp in length
* 41431 41530: gap of 100 bp
* 41531 42207: contig of 677 bp in length
* 42208 42307: gap of 100 bp
* 42308 42990: contig of 683 bp in length
* 42991 43090: gap of 100 bp
* 43091 43733: contig of 643 bp in length
* 43734 43833: gap of 100 bp
* 43834 44550: contig of 717 bp in length
* 44551 44650: gap of 100 bp
* 44651 45375: contig of 725 bp in length
* 45376 45475: gap of 100 bp
* 45476 46180: contig of 705 bp in length
* 46181 46280: gap of 100 bp
* 46281 46981: contig of 701 bp in length
* 46982 47081: gap of 100 bp
* 47082 47778: contig of 697 bp in length
* 47779 47878: gap of 100 bp
* 47879 48561: contig of 683 bp in length
* 48562 48661: gap of 100 bp
* 48662 49359: contig of 698 bp in length
* 49360 49459: gap of 100 bp
* 49460 50140: contig of 681 bp in length
* 50141 50240: gap of 100 bp
* 50241 50927: contig of 687 bp in length
* 50928 51027: gap of 100 bp
* 51028 51703: contig of 676 bp in length
* 51704 51803: gap of 100 bp
* 51804 52490: contig of 687 bp in length
* 52491 52590: gap of 100 bp
* 52591 53290: contig of 700 bp in length
* 53291 53390: gap of 100 bp
* 53391 54113: contig of 723 bp in length
* 54114 54213: gap of 100 bp
* 54214 54910: contig of 697 bp in length

```

* 54911 55010: gap of 100 bp

Query Match 50.08; Score 26; DB 2; Length 65370;
Best Local Similarity 70.08; Pred. No. 96;
Matches 35; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 1 aataaagattccattggataagaacgaatctgttacttgcgtcagt 50
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 64983 AATAAATATTCATTTTATATAAAGAAATGTTGCTTCAGTTTCTGT 65032

RESULT 10
ALI139381
LOCUS 116021 bp DNA linear PRI 12-OCT-2001
DEFINITION Human DNA sequence from clone RP11-74A12 on chromosome
13q31.2-32.2, complete sequence.
ACCESSION ALI139381
VERSION ALI139381.24 GI:161116424
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 116021)
Direct Submission
Submitted (11-OCT-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Oct 12, 2001 this sequence version replaced gi:15990615.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em.: EMBL; Sw.:
SWISSPROT; Tr.: TrEMBL; Wp.: WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
This sequence was generated from part of bacterial clone contigs of human
chromosome 13, constructed by the Sanger Centre Chromosome 13
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr13
RP11-74A12 is from the library RPC1-11.1 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACE3.6
IMPORTANT: This sequence is not the entire insert of clone
RP11-74A12 it may be shorter because we sequence overlapping
sections only once, except for a short overlap.
The true left end of clone RP11-46412 is at 114022 in this
sequence. The true right end of clone RP11-140119 is at 2000 in
this sequence.

FEATURES
source
Location/Qualifiers
1..116021
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/map="q31.2-32.2"
/clone="RP11-74A12"
/clone_lib="RPC1-11.1"
complement(88879..88909)
/note="Single clone region. Assembly confirmed by
restriction digest data."

misc_feature
Query Match 49.68; Score 25.8; DB 9; Length 116021;
Best Local Similarity 73.38; Pred. No. 1.1e+02;
Matches 33; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 1 aataaagattccattggataagaacgaatctgttacttgcgtt 45
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 103282 AGAAAATATTCACCTTGATGAGATCAAGCTGTGTTATTCCTT 103326

RESULT 11
AC008771
LOCUS 123169 bp DNA linear PRI 28-FEB-2001
DEFINITION Homo sapiens chromosome 5 clone CTD-2015H6, complete sequence.
ACCESSION AC008771
VERSION AC008771.4 GI:13162500
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 123169)
Direct Submission
Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 123169)
Direct Submission
Submitted (28-FEB-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
On Feb 28, 2001 this sequence version replaced gi:7709299.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.3% of Sequence;
Estimated Total Number of Errors is 0.9.
STS Content:
SHGC-84992 G53847
WI-15051 G23648
WI-17790 G24283.
FEATURES
source
Location/Qualifiers
1..123169
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTD-2015H6"
BASE COUNT 35484 a 23783 c 24111 g 39791 t
ORIGIN

Query Match 49.68; Score 25.8; DB 9; Length 123169;
Best Local Similarity 73.38; Pred. No. 1.1e+02;
Matches 33; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 1 aataaagattccattggataagaacgaatctgttacttgcgtt 45
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 79746 AGTAAAGATCCCTTGGATGATATATACATGTTATTTTGATT 79790

RESULT 12
AC018764/c

```

```

LOCUS       AC018764               126054 bp    DNA    linear    PRI 09-NOV-2000
DEFINITION  Homo sapiens chromosome 5 clone CTD-2327L5, complete sequence.
ACCESSION   AC018764
VERSION     AC018764.6  GI:11128366
KEYWORDS    HTG.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 126054)
AUTHORS    DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE      Direct Submission
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 126054)
AUTHORS    DOE Joint Genome Institute.
TITLE      Direct Submission
JOURNAL     Submitted (19-DEC-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3 (bases 1 to 126054)
AUTHORS    DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE      Direct Submission
JOURNAL     Submitted (01-JUL-2000) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
REFERENCE   4 (bases 1 to 126054)
AUTHORS    DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE      Direct Submission
JOURNAL     Submitted (09-NOV-2000) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Nov 9, 2000 this sequence version replaced gi:8886981.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www.shgc.stanford.edu
            Quality: Phrap Quality >=40 99.5% of Sequence;
            Estimated Total Number of Errors is 0.5.
            STS Content:
            WI-17790 G24283
            WI-15051 G23648
            Bases 89328 to 90656 excised as IS:Tn10.

FEATURES             source
     source           1..126054
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CTD-2327L5"
BASE COUNT  38749 a 24943 c 24689 g 37673 t
ORIGIN
Query Match      49.6%; Score 25.8; DB 9; Length 126054;
Best Local Similarity 73.3%; Pred. No. 1.1e+02;
Matches 33; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy  1 aataaaagattccatttgataaagaacgaactcttactttgctt 45
    | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 101816 AGTAAAGATTCCCTTGGTCATATACGTGATTATTTTGATT 101772

RESULT 13
AC090825/c
LOCUS       AC090825               143867 bp    DNA    linear    HTG 18-NOV-2001
DEFINITION  Homo sapiens chromosome 15 clone CTD-2054N24 map 15, WORKING DRAFT
            SEQUENCE, 3 unordered pieces.
ACCESSION   AC090825
VERSION     AC090825.3  GI:16974176
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 143867)
AUTHORS    Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE      Homo sapiens chromosome 15, clone CTD-2054N24

```

JOURNAL
REFERENCE
AUTHORS

Unpublished
2 (bases 1 to 143867)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S.,
Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B., Brown, A.,
Camrata, J., Campopiano, A., Choepel, Y., Colangelo, M., Collins, S.,
Collamore, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S.,
Dodgson, S., Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J.,
Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Karatas, A., LaRocque, K., Lamazares, R., Landers, T.,
Lehoczky, J., Levine, R., Liu, G., MacLean, C., Macdonald, P.,
Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K.,
McPheeters, R., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V.,
Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H.,
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R.,
Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M.,
Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S., Severy, P.,
Sougnuez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A.,
Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J.,
Zembek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL

Direct Submission
Submitted (10-MAR-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Nov 18, 2001 this sequence version replaced gi:14336507.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

COMMENT

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L12840
Center clone name: 2054_N_24
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 143329 bases at least Q40
Consensus quality: 143488 bases at least Q30
Consensus quality: 143566 bases at least Q20
Insert size: 172000; agarose-fp
Insert size: 143667; sum-of-contigs
Quality coverage: 10.5 in Q20 bases; agarose-fp
Quality coverage: 12.5 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 24723: contig of 24723 bp in length
* 24724 24823: gap of 100 bp
* 24824 46565: contig of 21742 bp in length
* 46566 46665: gap of 100 bp
* 46666 143867: contig of 97202 bp in length.
* Location/Qualifiers
1. .143867
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/map="15"
/clone="CTD-2054N24"
/clone_lib="CITD Human BAC"
1. .24723
/note="assembly_fragment"

FEATURES
source

misc_feature
1. .24723
/note="assembly_fragment"

misc_feature 24824 .46565 /note="assembly_fragment"

misc_feature 46666 .143867 /note="assembly_fragment"

BASE COUNT 45409 a 27723 c 27453 g 43082 t 200 others

ORIGIN

Query Match 49.6%; Score 25.8; DB 2; Length 143867;
Best Local Similarity 73.3%; Pred. No. 1e+02;
Matches 33; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 1 ataaagaattccattggataagaacgaatctgttacttctgtt 45

Db 94755 AAAAAATTTTCTGCTGGATGAGACGATCTGCTACTGTGTTT 94711

RESULT 14
AC096094 176476 bp DNA linear HTG 20-DEC-2001
LOCUS Rattus norvegicus clone CH230-24H1, *** SEQUENCING IN PROGRESS ***,
DEFINITION 62 unordered pieces.
AC096094
AC096094.2 GI:17943777
VERSION HTG; HTGS-PHASE1.
KEYWORDS Rattus norvegicus
SOURCE Norway rat.
ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 176476)

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
Alsbrooks,S.L., Amarantunga,H.C., Are,J.R., Banks,T., Barbara,J.,
Benton,K., Blimie,K., Blomquist,K., Bonnin,D., Bouck,J.,
Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhury,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dedrich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
Hollins,B., Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Loulsegh,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P., Meador,M.,
Mel,G., Metzger,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenwo,S.,
Ogih,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojebokan,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshbari,N.,
Sisson,I., Sodergren,E., Sonaik,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Taber,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.

Direct Submission

Unpublished

2 (bases 1 to 176476)

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE
JOURNAL

COMMENT

Worley,K.C.

Direct Submission

Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Dec 20, 2001 this sequence version replaced gi:15627714.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GEKO

Center clone name: CH230-24H1

----- Summary Statistics

Assembly program: Phrap; version 0.990329First call to

findPhrapList

Consensus quality: 140728 bases at least Q40
Consensus quality: 150414 bases at least Q30
Consensus quality: 157122 bases at least Q20
Estimated insert size: 149244; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 2.3x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
consists of 62 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
be preserved.

1
4911: contig of 4911 bp in length
4912
5011: gap of unknown length
5012
10561: contig of 5550 bp in length
10562
10661: gap of unknown length
10662
13934: contig of 3273 bp in length
13935
14034: gap of unknown length
20395
20595: contig of 6561 bp in length
20596
20596: gap of unknown length
20696
27180: contig of 6485 bp in length
27181
27281: gap of unknown length
33208: contig of 5928 bp in length
33209
37842: gap of unknown length
37843
37942: gap of unknown length
41481: contig of 3539 bp in length
41482
41581: gap of unknown length
41582
47156: contig of 5575 bp in length
47157
47256: gap of unknown length
47257
51202: contig of 4264 bp in length
51521
51620: gap of unknown length
51621
54846: contig of 3226 bp in length
54847
54946: gap of unknown length
54947
58569: contig of 3623 bp in length
58570
58669: gap of unknown length
61873: contig of 3204 bp in length
61874
61973: gap of unknown length
61974
66179: contig of 4206 bp in length
66180
66279: gap of unknown length
66280
71150: contig of 4871 bp in length
71151
71250: gap of unknown length
71251
75339: contig of 4089 bp in length
75340
75439: gap of unknown length
79077: contig of 3638 bp in length
79078
79177: gap of unknown length
79178
82408: contig of 3231 bp in length
82409
82508: gap of unknown length
87001: contig of 4493 bp in length
87002
87101: gap of unknown length
87102
91317: contig of 4216 bp in length

* This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 48313: contig of 48313 bp in length
 * 48314 48413: gap of 100 bp
 * 48414 50035: contig of 1622 bp in length
 * 50036 50135: gap of 100 bp
 * 50136 54638: contig of 4503 bp in length
 * 54639 54738: gap of 100 bp
 * 54739 58868: contig of 4130 bp in length
 * 58869 58968: gap of 100 bp
 * 58969 64007: contig of 5039 bp in length
 * 64008 64107: gap of 100 bp
 * 64108 71220: contig of 7113 bp in length
 * 71221 71320: gap of 100 bp
 * 71321 76136: contig of 4816 bp in length
 * 76137 76236: gap of 100 bp
 * 76237 82769: contig of 6533 bp in length
 * 82770 82869: gap of 100 bp
 * 82870 104915: contig of 22046 bp in length
 * 104916 105015: gap of 100 bp
 * 105016 119090: contig of 14075 bp in length
 * 119091 119190: gap of 100 bp
 * 119191 167682: contig of 48492 bp in length
 * 167683 167782: gap of 100 bp
 * 167783 179762: contig of 11980 bp in length.

FEATURES

source 1. .179762
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="RP11-115014"
 /clone="RP11-115014" Human Male BAC"
 misc_feature 1. .48313
 /note="assembly_fragment"
 clone_end:SP6
 vector_side:left
 48414..50035
 /note="assembly_fragment"
 misc_feature 50136..54638
 /note="assembly_fragment"
 misc_feature 54739..58868
 /note="assembly_fragment"
 misc_feature 58969..64007
 /note="assembly_fragment"
 misc_feature 64108..71220
 /note="assembly_fragment"
 misc_feature 71321..76136
 /note="assembly_fragment"
 misc_feature 76237..82769
 /note="assembly_fragment"
 misc_feature 82870..104915
 /note="assembly_fragment"
 misc_feature 105016..119090
 /note="assembly_fragment"
 misc_feature 119191..167682
 /note="assembly_fragment"
 misc_feature 167783..179762
 /note="assembly_fragment"
 clone_end:T7
 vector_side:right
 BASE COUNT 52681 a 37316 c 36703 g 51962 t 1100 others
 ORIGIN

Query Match 49.6%; Score 25.8; DB 2; Length 179762;
 Best Local Similarity 73.3%; Pred. No. le+02;
 Matches 33; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 1 aataaaagattccattgataagaacgaactgttacttgcctt 45
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 159094 AAAAAATTTTCTCTGGTGGATGAGAACGGATCTGGTACTGTGTTT 159138

Search completed: June 19, 2002, 15:50:29
 Job time: 13887 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2002, 17:14:11 ; Search time 3798.65 Seconds
(without alignments)
23.503 Million cell updates/sec

Title: US-09-462-955B-1_COPY_734_785

Perfect score: 52

Sequence: 1 aataaagattccattgga.....tggtactttgttcgagtc 52

Scoring table:

IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_032802.*

1: /net/abss06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1980.DAT.*
2: /net/abss06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1981.DAT.*
3: /net/abss06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1982.DAT.*
4: /net/abss06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1983.DAT.*
5: /net/abss06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1984.DAT.*
6: /net/abss06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1985.DAT.*
7: /net/abss06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1986.DAT.*
8: /net/abss06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1987.DAT.*
9: /net/abss06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1988.DAT.*
10: /net/abss06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1989.DAT.*
11: /net/abss06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1990.DAT.*
12: /net/abss06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1991.DAT.*
13: /net/abss06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1992.DAT.*
14: /net/abss06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1993.DAT.*
15: /net/abss06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1994.DAT.*
16: /net/abss06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1995.DAT.*
17: /net/abss06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1996.DAT.*
18: /net/abss06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1997.DAT.*
19: /net/abss06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1998.DAT.*
20: /net/abss06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1999.DAT.*
21: /net/abss06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA2000.DAT.*
22: /net/abss06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA2001A.DAT.*
23: /net/abss06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA2001B.DAT.*
24: /net/abss06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	26.4	50.8	243	21	Promoter from inte
C 2	25.4	48.8	6198	22	Human ovarian and
C 3	25.4	48.8	6198	22	Human reproductive
C 4	25	48.1	811	22	Human CDNA clone (
C 5	24.6	47.3	18765	24	Human glutamate re
C 6	24.6	47.3	1082138	21	Arabidopsis thalia
C 7	24.4	46.9	2763	22	Human CDNA sequenc
C 8	24.2	46.5	1173	22	Plasmid pBNpnt ins
C 9	24.2	46.5	15366	22	Genomic sequence #

C 10	23.4	45.0	575	22	AAH70492	Human cervical can
C 11	23.4	45.0	1113	22	AAH73180	Human cervical can
C 12	23.4	45.0	20835	22	AAK86765	Human immune/haema
C 13	23.4	45.0	25806	22	AAK86766	Human immune/haema
C 14	23.4	45.0	580073	18	AAH58840	Mycoplasma genital
C 15	23.2	44.6	2756	22	AAH14861	Human CDNA sequenc
C 16	23.2	44.6	6305	16	AAQ96127	Tobacco leaf antif
C 17	23	44.2	403	24	AAH17049	Human protein kina
C 18	23	44.2	578	22	AAH16113	Human colon cancer
C 19	23	44.2	732	21	AAH97370	Human colorectal c
C 20	23	44.2	1079	22	AAH60801	Human cancer agent
C 21	23	44.2	1931	21	AAH16149	Human prostate can
C 22	22.8	43.8	338	21	AAH01250	Human secreted pro
C 23	22.8	43.8	459	22	AAK83834	Human immune/haema
C 24	22.8	43.8	459	22	AAK83835	Human immune/haema
C 25	22.8	43.8	546	23	AAH78220	DNA encoding novel
C 26	22.8	43.8	790	22	AAH08051	Human CDNA clone (
C 27	22.8	43.8	1164	22	AAH161292	Human polynucleoti
C 28	22.8	43.8	1238	22	AAH192942	Human polynucleoti
C 29	22.8	43.8	1473	22	AAH195056	Human polynucleoti
C 30	22.8	43.8	2459	23	ABL26096	Drosophila melanog
C 31	22.8	43.8	2506	22	AAH15133	Human CDNA sequenc
C 32	22.8	43.8	3206	22	AAH58337	Human GTP-binding
C 33	22.8	43.8	8043	16	AAH58524	Human protein tyro
C 34	22.8	43.8	8119	21	AAH61840	DNA encoding a hum
C 35	22.8	43.8	16449	23	ABL25556	Drosophila melanog
C 36	22.8	43.8	465237	24	ABA90193	Human oestrogen re
C 37	22.6	43.5	429	22	AAH53000	S. epidermidis ope
C 38	22.6	43.5	604	20	AAH10665	cDNA encoding a hu
C 39	22.6	43.5	2718	22	AAH160849	Human polynucleoti
C 40	22.6	43.5	3038	20	AAH234199	Human PRO541 nucle
C 41	22.6	43.5	3038	21	AAH78550	Human PRO541 (UNQ3
C 42	22.6	43.5	3038	22	AAH45959	Human DNA encoding
C 43	22.6	43.5	3145	22	AAH54834	S. epidermidis gen
C 44	22.6	43.5	3771	22	AAH54702	S. epidermidis gen
C 45	22.6	43.5	6850	24	AAH61365	Human gene regulat

ALIGNMENTS

RESULT	1
AAA96835/c	
ID	AAA96835 standard; DNA; 243 BP.
XX	
AC	AAA96835;
XX	
DT	19-FEB-2001 (first entry)
XX	
DE	Promoter from intergenic region of Commelina yellow mottle virus.
DE	
DE	Promoter; intergenic region; Commelina yellow mottle virus;
KW	chimeric expression promoter; plant vascular expression promoter;
KW	plant green tissue expression promoter; Cassava vein mosaic virus;
KW	transgenic plant; ss.
XX	
OS	Commelina yellow mottle virus.
XX	
PN	WO200058485-A1.
PD	
PD	05-OCT-2000.
XX	
XX	29-MAR-2000; 2000WO-IB00370.
PF	
PR	29-MAR-1999; 99FR-0003925.
XX	
PA	(MERI-) MERISTEM THERAPEUTICS.
XX	
PI	Rance I, Gruber V, Theisen M;
XX	
XX	WPI; 2000-647238/62.
DR	
XX	
PT	Chimeric expression promoter for transgenic plant production, comprises


```
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-02559678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-465570/50.
XX
XX Isolated nucleic acid molecule encoding a reproductive system antigen
XX is used in preventing, treating or ameliorating a medical condition -
XX
XX Disclosure; SEQ ID NO 9612; 1297pp + Sequence Listing; English.
XX
XX The present invention provides the protein and coding sequences of a
XX number of human reproductive system related antigens. These can be used
XX in the prevention and treatment of reproductive system disorders,
XX including cancer. The present sequence is a genomic sequence encoding a
XX protein of the invention.
XX
XX Sequence 6198 BP; 1979 A; 1100 C; 1215 G; 1904 T; 0 other;
SQ
Query Match 48.8%; Score 25.4; DB 22; Length 6198;
Best Local Similarity 74.4%; Pred. No. 11;
Matches 32; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
QY 6 aagattccattgttgataagaacgaatctgttactttgttgca 48
Db 4704 atgacagcgtttgttgataaacaattctgttactttgtgatgca 4746
RESULT 4
AAH04923
ID AAH04923 standard; cDNA; 811 BP.
XX
XX AC AAH04923;
XX
XX 26-JUN-2001 (first entry)
XX
XX Human cDNA clone (5'-primer) SEQ ID NO:1758.
XX
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX Homo sapiens.
XX EP1074617-A2.
XX
```



```

PR 13-SEP-1999; 99US-0153584.
PR 17-SEP-1999; 99US-0154603.
PA (UYCH-) UNIV CHICAGO.
XX PI Preuss D, Copenhaver G, Keith K;
XX DR WPI; 2000-587529/55.
XX PT Recombinant DNA construct comprising a plant centromere, useful for
XX PT producing stably inherited microsome which can serve as vectors for
XX PT the construction of transgenic plant and animal cells -
XX PS Claim 68; Page 977-1388; 1449pp; English.
XX CC The present invention relates to a recombinant DNA construct of a plant
XX CC (Arabidopsis thaliana) centromere. The constructs are useful for
XX CC producing stably inherited microsome which can serve as vectors for
XX CC the construction of transgenic plant and animal cells expressing
XX CC selected proteins such as hormones, enzymes, interleukins, clotting
XX CC factors, cytokines, antibodies, and growth factors.
XX SQ Sequence 1082138 BP; 348775 A; 194404 C; 195515 G; 343444 T; 0 other;

Query Match 47.3%; Score 24.6; DB 21; Length 1082138;
Best Local Similarity 70.2%; Pred. No. 55;
Matches 33; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Qy 6 aagattccattgataagaacgaatctgttacttctgttcgagtc 52
||||| ||||| ||| ||||| ||||| ||||| ||||| |||||
Db 492337 AAGATAATATTAGAGAAATGAAGAATCTGAGACTTTGTTGTGGTTC 492291

RESULT 7
AAH17538
ID AAH17538 standard; cDNA; 2763 BP.
XX AC AAH17538;
XX DT 26-JUN-2001 (first entry)
XX DE Human cDNA sequence SEQ ID NO:17015.
XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX OS Homo sapiens.
XX PN EP1074617-A2.
XX PD 07-FEB-2001.
XX PF 28-JUL-2000; 2000EP-0116126.
XX PR 29-JUL-1999; 99JP-0248036.
XX PR 27-AUG-1999; 99JP-0300253.
XX PR 11-JAN-2000; 2000JP-0118776.
XX PR 02-MAY-2000; 2000JP-0183767.
XX PR 09-JUN-2000; 2000JP-0241899.
XX PA (HELI-) HELIX RES INST.
XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX DR WPI; 2001-318749/34.
XX PT Primer sets for synthesizing polynucleotides, particularly the 5602
XX PT full-length cDNAs defined in the specification, and for the detection
XX PT and/or diagnosis of the abnormality of the proteins encoded by the
XX PT full-length cDNAs -
XX PS Claim 8; SEQ ID 17015; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesizing 5602
XX full-length cDNAs defined in the specification. Where a primer set
XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX to the complementary strand of a polynucleotide which comprises one of
XX the 5602 nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises a 3'-end sequence, where the
XX oligonucleotide comprises at least 15 nucleotides and the combination of
XX the 5'-end sequence/3'-end sequence is selected from those defined in
XX the specification. The primer sets can be used in antisense therapy and
XX in gene therapy. The primers are useful for synthesizing polynucleotides,
XX particularly full-length cDNAs. The primers are also useful for the
XX detection and/or diagnosis of the abnormality of the proteins encoded by
XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
XX AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
XX represent oligonucleotides, all of which are used in the exemplification
XX of the present invention.
XX SQ Sequence 2763 BP; 823 A; 564 C; 635 G; 741 T; 0 other;

Query Match 46.9%; Score 24.4; DB 22; Length 2763;
Best Local Similarity 73.8%; Pred. No. 23;
Matches 31; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 8 gattccattgataagaacgaatctgttacttctgttcgag 49
||||| ||||| ||| ||||| ||||| ||||| ||||| |||||
Db 611 gttccattgtttaattcttacttctgttacttctgttcgag 652

RESULT 8
AAT84527/c
ID AAT84527 standard; DNA; 1173 BP.
XX AC AAT84527;
XX DT 02-DEC-1997 (first entry)
XX DE Plasmid pBNppt insert.
XX KW Subtilisin inhibitor; protease inhibitor; PSTI; human;
XX KW pancreatic secretory trypsin inhibitor; enzyme engineering;
XX KW protein engineering; detergent; ss.
XX OS Chimeric Bacillus subtilis strain BG2036;
XX OS Chimeric Bacillus amyloliquefaciens;
XX OS Chimeric synthetic.
XX FH Key Location/Qualifiers
XX FT Promoter 1..609
XX FT /*tag= a
XX FT /*note= "B. subtilis aprE promoter"
XX FT CDS 610..930
XX FT /*tag= b
XX FT /*note= "B. amyloliquefaciens subtilisin
XX FT gene signal peptide, prosequence and
XX FT terminator"
XX FT sig_peptide 610..699
XX FT /*tag= c
XX PN W09715670-A1.
XX XX 01-MAY-1997.
XX PD 25-OCT-1996; 96WO-US17153.
XX PF 25-OCT-1995; 95US-0548186.
XX PR

```



```

XX (ARRI-) ARRIS PHARM CORP.
XX PA
XX Estell DA, Fisher J, Hartman C, McGrath M, Schmidt B;
XX PI
XX WPI; 1997-259023/23.
XX DR P-PSDB; AAW26365.
XX DR
XX Modified subtilisin inhibitor - having altered reactive site
XX PT resulting in decreased immunogenicity, used in detergent
XX PT composition(s)
XX PS
XX Example 1; Fig 12; 69pp; English.
XX
XX This DNA sequence comprises a DNA insert in vector pBN3 comprising
XX the Bacillus subtilis aprE promoter and the bacillus
XX amyloliquefaciens subtilisin gene signal sequence, pro-sequence and
XX terminator region (see AAW26365). The vector, pBNpt, can be used
XX for the expression in B. subtilis of novel subtilisin inhibitors in
XX which a human standard mechanism inhibitor, such as pancreatic
XX secretory trypsin inhibitor, has at least one amino acid of the
XX reactive site substituted so that its dissociation constant for
XX inhibition of subtilisin is reduced by at least a factor of 100.
XX The novel subtilisin inhibitors (see AAW26362-64) combine the low
XX allergenicity of human standard mechanism inhibitors, which are not
XX specific for subtilisin, and the high affinity of non-human
XX subtilisin inhibitors such as Streptomyces subtilisin inhibitor and
XX turkey mucoid third domain protein. They form a complex with
XX subtilisin, so as to avoid problems of allergenicity in detergent
XX formulations, and can also be used in affinity purification and
XX (diagnostic) quantification of subtilisin.
XX
XX Sequence 1173 BP; 358 A; 249 C; 230 G; 336 T; 0 other;
XX
XX
XX Query Match 46.5%; Score 24.2; DB 18; Length 1173;
XX Best Local Similarity 78.4%; Pred. No. 23;
XX Matches 29; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
XX
XX Qy 13 catttgataagaacgaatctgttacttctgttcag 49
XX | ||||| || |||| ||||| |||| | ||||
XX Db 55 CGTTGGAAATCAGAGTCTGTATTATTGATAGCAG 19
XX
XX RESULT 9
XX AAS28623/c
XX ID AAS28623 standard; DNA; 15366 BP.
XX AC
XX AAS28623;
XX XX
XX 07-NOV-2001 (first entry)
XX XX
XX Genomic sequence #463 encoding for novel human respiratory antigen.
XX
XX Human; respiratory antigen; respiratory disorder; throat disorder;
XX lung disorder; nose disorder; lung cancer; gene therapy; cytostatic;
XX anti allergic; anti asthmatic; anti inflammatory; olfactory;
XX respiratory active; ds.
XX
XX OS
XX Homo sapiens.
XX
XX WO20015448-A1.
XX PN
XX 02-AUG-2001.
XX PD
XX
XX 17-JAN-2001; 2001WO-0501333.
XX PF
XX
XX 31-JAN-2000; 2000US-0179065.
XX PR
XX 04-FEB-2000; 2000US-0180628.
XX PR
XX 24-FEB-2000; 2000US-0184564.
XX PR
XX 02-MAR-2000; 2000US-0186350.
XX PR
XX 16-MAR-2000; 2000US-0189874.
XX PR
XX 17-MAR-2000; 2000US-0190076.
XX PR
XX
XX 18-APR-2000; 2000US-0198123.
XX PR
XX 19-MAY-2000; 2000US-0205515.
XX PR
XX 07-JUN-2000; 2000US-0209467.
XX PR
XX 28-JUN-2000; 2000US-0214886.
XX PR
XX 30-JUN-2000; 2000US-0215135.
XX PR
XX 07-JUL-2000; 2000US-0216647.
XX PR
XX 07-JUL-2000; 2000US-0216880.
XX PR
XX 11-JUL-2000; 2000US-0217487.
XX PR
XX 11-JUL-2000; 2000US-0217496.
XX PR
XX 14-JUL-2000; 2000US-0218290.
XX PR
XX 26-JUL-2000; 2000US-0220963.
XX PR
XX 26-JUL-2000; 2000US-0220964.
XX PR
XX 14-AUG-2000; 2000US-0224518.
XX PR
XX 14-AUG-2000; 2000US-0224519.
XX PR
XX 14-AUG-2000; 2000US-0225213.
XX PR
XX 14-AUG-2000; 2000US-0225214.
XX PR
XX 14-AUG-2000; 2000US-0225266.
XX PR
XX 14-AUG-2000; 2000US-0225267.
XX PR
XX 14-AUG-2000; 2000US-0225268.
XX PR
XX 14-AUG-2000; 2000US-0225270.
XX PR
XX 14-AUG-2000; 2000US-0225447.
XX PR
XX 14-AUG-2000; 2000US-0225757.
XX PR
XX 14-AUG-2000; 2000US-0225758.
XX PR
XX 14-AUG-2000; 2000US-0225759.
XX PR
XX 18-AUG-2000; 2000US-0226279.
XX PR
XX 22-AUG-2000; 2000US-0226681.
XX PR
XX 22-AUG-2000; 2000US-0226868.
XX PR
XX 22-AUG-2000; 2000US-0227182.
XX PR
XX 23-AUG-2000; 2000US-0227009.
XX PR
XX 30-AUG-2000; 2000US-0228924.
XX PR
XX 01-SEP-2000; 2000US-0229287.
XX PR
XX 01-SEP-2000; 2000US-0229343.
XX PR
XX 01-SEP-2000; 2000US-0229344.
XX PR
XX 01-SEP-2000; 2000US-0229345.
XX PR
XX 05-SEP-2000; 2000US-0229509.
XX PR
XX 05-SEP-2000; 2000US-0229513.
XX PR
XX 06-SEP-2000; 2000US-0230437.
XX PR
XX 06-SEP-2000; 2000US-0230438.
XX PR
XX 08-SEP-2000; 2000US-0231242.
XX PR
XX 08-SEP-2000; 2000US-0231243.
XX PR
XX 08-SEP-2000; 2000US-0231244.
XX PR
XX 08-SEP-2000; 2000US-0231413.
XX PR
XX 08-SEP-2000; 2000US-0231414.
XX PR
XX 08-SEP-2000; 2000US-0232080.
XX PR
XX 08-SEP-2000; 2000US-0232081.
XX PR
XX 12-SEP-2000; 2000US-0231968.
XX PR
XX 14-SEP-2000; 2000US-0232397.
XX PR
XX 14-SEP-2000; 2000US-0232398.
XX PR
XX 14-SEP-2000; 2000US-0232399.
XX PR
XX 14-SEP-2000; 2000US-0232400.
XX PR
XX 14-SEP-2000; 2000US-0232401.
XX PR
XX 14-SEP-2000; 2000US-0233063.
XX PR
XX 14-SEP-2000; 2000US-0233064.
XX PR
XX 14-SEP-2000; 2000US-0233065.
XX PR
XX 21-SEP-2000; 2000US-0234223.
XX PR
XX 21-SEP-2000; 2000US-0234274.
XX PR
XX 25-SEP-2000; 2000US-0234997.
XX PR
XX 25-SEP-2000; 2000US-0234998.
XX PR
XX 25-SEP-2000; 2000US-0235484.
XX PR
XX 26-SEP-2000; 2000US-0235484.
XX PR
XX 27-SEP-2000; 2000US-0235834.
XX PR
XX 27-SEP-2000; 2000US-0235836.
XX PR
XX 29-SEP-2000; 2000US-0236327.
XX PR
XX 29-SEP-2000; 2000US-0236367.
XX PR
XX 29-SEP-2000; 2000US-0236368.
XX PR
XX 29-SEP-2000; 2000US-0236369.
XX PR
XX 29-SEP-2000; 2000US-0236370.
XX PR
XX 02-OCT-2000; 2000US-0236802.
XX PR
XX 02-OCT-2000; 2000US-0237037.
XX PR
XX 02-OCT-2000; 2000US-0237038.
XX PR
XX 02-OCT-2000; 2000US-0237039.
XX PR
XX 02-OCT-2000; 2000US-0237040.
XX PR
XX 13-OCT-2000; 2000US-0239935.
XX PR
XX 13-OCT-2000; 2000US-0239937.
XX PR

```

20-OCT-2000; 2000US-0240960.
 20-OCT-2000; 2000US-0241221.
 20-OCT-2000; 2000US-0241785.
 20-OCT-2000; 2000US-0241786.
 20-OCT-2000; 2000US-0241787.
 20-OCT-2000; 2000US-0241808.
 20-OCT-2000; 2000US-0241809.
 20-OCT-2000; 2000US-0241826.
 01-NOV-2000; 2000US-0244617.
 08-NOV-2000; 2000US-0246474.
 08-NOV-2000; 2000US-0246475.
 08-NOV-2000; 2000US-0246476.
 08-NOV-2000; 2000US-0246477.
 08-NOV-2000; 2000US-0246523.
 08-NOV-2000; 2000US-0246524.
 08-NOV-2000; 2000US-0246525.
 08-NOV-2000; 2000US-0246526.
 08-NOV-2000; 2000US-0246527.
 08-NOV-2000; 2000US-0246528.
 08-NOV-2000; 2000US-0246532.
 08-NOV-2000; 2000US-0246609.
 08-NOV-2000; 2000US-0246610.
 08-NOV-2000; 2000US-0246611.
 08-NOV-2000; 2000US-0246613.
 17-NOV-2000; 2000US-0249207.
 17-NOV-2000; 2000US-0249208.
 17-NOV-2000; 2000US-0249209.
 17-NOV-2000; 2000US-0249210.
 17-NOV-2000; 2000US-0249211.
 17-NOV-2000; 2000US-0249212.
 17-NOV-2000; 2000US-0249213.
 17-NOV-2000; 2000US-0249214.
 17-NOV-2000; 2000US-0249215.
 17-NOV-2000; 2000US-0249216.
 17-NOV-2000; 2000US-0249217.
 17-NOV-2000; 2000US-0249218.
 17-NOV-2000; 2000US-0249244.
 17-NOV-2000; 2000US-0249245.
 17-NOV-2000; 2000US-0249264.
 17-NOV-2000; 2000US-0249265.
 17-NOV-2000; 2000US-0249297.
 17-NOV-2000; 2000US-0249299.
 17-NOV-2000; 2000US-0249300.
 01-DEC-2000; 2000US-0250160.
 01-DEC-2000; 2000US-0250391.
 05-DEC-2000; 2000US-0251030.
 05-DEC-2000; 2000US-0251988.
 05-DEC-2000; 2000US-0256719.
 06-DEC-2000; 2000US-0251479.
 08-DEC-2000; 2000US-0251856.
 08-DEC-2000; 2000US-0251868.
 08-DEC-2000; 2000US-0251869.
 08-DEC-2000; 2000US-0251989.
 08-DEC-2000; 2000US-0251990.
 11-DEC-2000; 2000US-0254097.
 05-JAN-2001; 2001US-0259678.
 (HUMA-) HUMAN GENOME SCI INC.
 Rosen CA, Barash SC, Ruben SM;
 WPI; 2001-476224/51.
 Isolated polypeptide for treating, preventing and/or prognosing
 disorders related to the respiratory system including respiratory
 cancers and also for testing and detection e.g. diagnosis -
 Disclosure: SED ID No 1057; 546pp; English.
 The present invention relates to the isolation of novel human
 respiratory antigens (AAU17685-AAU17975), and cDNA and genomic
 sequences encoding for these polypeptides. The sequences of the

invention are useful for preventing, treating and/or prognosing
 disorders related to the respiratory system including throat
 disorders (e.g. vocal cord paralysis, tonsillitis, and laryngitis),
 lung disorders e.g. pneumonia, allergic disorders e.g. asthma,
 pleurisy, cystic fibrosis, emphysema, nose disorders and cancers of
 the respiratory tissues e.g. lung cancer. The polynucleotide sequences
 of the invention are useful in gene therapy and antisense therapy.
 CC AAS28161-AAS28764 represent genomic sequences encoding for novel
 CC human respiratory antigens.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 15366 BP; 4516 A; 2790 C; 3055 G; 5005 T; 0 other;

Query Match 46.5%; Score 24.2; DB 22; Length 15366;
 Best Local Similarity 71.1%; Pred. No. 36;
 Matches 32; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 2 ataaagattccatttggaataacgaatctgttacttctgttg 46
 ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 15172 ACAGAAATTTCAATGTAATAAATAACGAAGCTGTACCTTGTCTGG 15128

RESULT 10
 AAH70492/c
 ID AAH70492 standard; cDNA: 575 BP.
 XX
 AC AAH70492;
 XX
 DT 19-SEP-2001 (first entry)
 XX
 DE Human cervical cancer marker nucleic acid 1766.
 XX
 KW Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200142467-A2.
 XX
 PD 14-JUN-2001.
 XX
 PF 08-DEC-2000; 2000WO-US33312.
 XX
 PR 08-DEC-1999; 99US-0169681.
 PR 21-DEC-1999; 99US-0171350.
 PR 14-MAR-2000; 2000US-0189315.
 PR 12-MAY-2000; 2000US-0203791.
 PR 09-JUN-2000; 2000US-0210600.
 PR 21-JUL-2000; 2000US-0220114.
 XX
 PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 XX
 PI Schlegel R, Deeds J, Berger A, Zhao X;
 XX
 XX WPI; 2001-375006/39.
 XX
 PT New isolated nucleic acid for diagnosing and treating cervical cancer
 PT and for assessing and detecting compounds for treating the cancer -
 XX
 PS Claim 1; Page 383; 1051pp; English.
 XX
 CC The invention relates to novel genes (AAH68727-AAH73383) associated with
 CC cervical cancer with cytostatic activity. The nucleic acids and encoded
 CC polypeptides are useful; to assess if a patient is afflicted with
 CC cervical cancer or has a pre-malignant condition; to monitor the
 CC progression of cervical cancer or a premalignant condition in a patient;
 CC and to select and/or assess the efficacy of a compound or therapy for
 CC inhibiting cervical cancer in a patient. The nucleic acids may also be
 CC useful for gene therapy.
 XX
 SQ Sequence 575 BP; 190 A; 85 C; 119 G; 181 T; 0 other;

Query Match 45.0%; Score 23.4; DB 22; Length 575;
Best Local Similarity 81.8%; Pred. No. 39;
Matches 27; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 10 ttccatttgataagaacgaatctgttactttg 42
| | ||||| |||| ||||| |||||
Db 392 TCCTTTTGGATAAAACGATATCTGTGCTTG 360

RESULT 11
AAH73180/c
ID AAH73180 standard; cDNA; 1113 BP.

XX AC

XX AAH73180;

DT 19-SEP-2001 (first entry)

XX Human cervical cancer marker nucleic acid 4454.

DE Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.

XX KW

XX OS

XX Homo sapiens.

XX WO200142467-A2.

PN 14-JUN-2001.

XX PD

XX PF 08-DEC-2000; 2000WO-US33312.

XX 08-DEC-1999; 99US-0169681.

XX 21-DEC-1999; 99US-0171350.

XX 14-MAR-2000; 2000US-0189315.

XX 12-MAY-2000; 2000US-0203791.

XX 09-JUN-2000; 2000US-0210600.

XX 21-JUL-2000; 2000US-0220114.

XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX Schlegel R, Deeds J, Berger A, Zhao X;

PI WPI; 2001-375006/39.

XX New isolated nucleic acid for diagnosing and treating cervical cancer

PT and for assessing and detecting compounds for treating the cancer -

XX Claim 1; Page 974; 1051pp; English.

XX The invention relates to novel genes (AAH68727-AAH73383) associated with
CC cervical cancer with cytostatic activity. The nucleic acids and encoded
CC polypeptides are useful: to assess if a patient is afflicted with
CC cervical cancer or has a pre-malignant condition; to monitor the
CC progression of cervical cancer or a premalignant condition in a patient;
CC and to select and/or assess the efficacy of a compound or therapy for
CC inhibiting cervical cancer in a patient. The nucleic acids may also be
CC useful for gene therapy.

XX Sequence 1113 BP; 358 A; 172 C; 223 G; 347 T; 13 other;

Query Match 45.0%; Score 23.4; DB 22; Length 1113;
Best Local Similarity 81.8%; Pred. No. 44;
Matches 27; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 10 ttccatttgataagaacgaatctgttactttg 42
| | ||||| |||| ||||| |||||
Db 774 TCCTTTTGGATAAAACGATATCTGTGCTTG 742

RESULT 12

AAK86765/c

ID AAK86765 standard; DNA; 20835 BP.

XX AC AAK86765;
XX 07-NOV-2001 (first entry)
DT Human immune/haematopoietic antigen genomic sequence SEQ ID NO:41577.
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX cytostatic; gene therapy; vaccine; metastasis; ds.
XX Homo sapiens.
XX WO200157182-A2.
XX 09-AUG-2001.
XX 17-JAN-2001; 2001WO-US01354.
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184664.
XX 02-MAR-2000; 2000US-0186350.
XX 16-MAR-2000; 2000US-0189874.
XX 17-MAR-2000; 2000US-0190076.
XX 18-APR-2000; 2000US-0198123.
XX 19-MAY-2000; 2000US-0205515.
XX 07-JUN-2000; 2000US-0209467.
XX 28-JUN-2000; 2000US-0214886.
XX 30-JUN-2000; 2000US-0215135.
XX 07-JUL-2000; 2000US-0216647.
XX 07-JUL-2000; 2000US-0216880.
XX 11-JUL-2000; 2000US-0217487.
XX 11-JUL-2000; 2000US-0217496.
XX 14-JUL-2000; 2000US-0218290.
XX 26-JUL-2000; 2000US-0220963.
XX 26-JUL-2000; 2000US-0220964.
XX 14-AUG-2000; 2000US-0224518.
XX 14-AUG-2000; 2000US-0224519.
XX 14-AUG-2000; 2000US-0225213.
XX 14-AUG-2000; 2000US-0225214.
XX 14-AUG-2000; 2000US-0225266.
XX 14-AUG-2000; 2000US-0225267.
XX 14-AUG-2000; 2000US-0225268.
XX 14-AUG-2000; 2000US-0225270.
XX 14-AUG-2000; 2000US-0225447.
XX 14-AUG-2000; 2000US-0225757.
XX 14-AUG-2000; 2000US-0225758.
XX 14-AUG-2000; 2000US-0225759.
XX 18-AUG-2000; 2000US-0226279.
XX 22-AUG-2000; 2000US-0226681.
XX 22-AUG-2000; 2000US-0226868.
XX 22-AUG-2000; 2000US-0227182.
XX 30-AUG-2000; 2000US-0227009.
XX 01-SEP-2000; 2000US-0228924.
XX 01-SEP-2000; 2000US-0229287.
XX 01-SEP-2000; 2000US-0229343.
XX 01-SEP-2000; 2000US-0229344.
XX 01-SEP-2000; 2000US-0229345.
XX 05-SEP-2000; 2000US-0229509.
XX 05-SEP-2000; 2000US-0229513.
XX 06-SEP-2000; 2000US-0230437.
XX 06-SEP-2000; 2000US-0230438.
XX 08-SEP-2000; 2000US-0231242.
XX 08-SEP-2000; 2000US-0231243.
XX 08-SEP-2000; 2000US-0231244.
XX 08-SEP-2000; 2000US-0231413.
XX 08-SEP-2000; 2000US-0231414.
XX 08-SEP-2000; 2000US-0232080.
XX 08-SEP-2000; 2000US-0232081.
XX 12-SEP-2000; 2000US-0232081.
XX 14-SEP-2000; 2000US-0231968.
XX 14-SEP-2000; 2000US-0232397.
XX 14-SEP-2000; 2000US-0232398.
XX 14-SEP-2000; 2000US-0232399.

```
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234224.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0234998.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-024617.
PR 08-NOV-2000; 2000US-0246174.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251890.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
PR (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-483426/52.
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
XX useful for preventing, diagnosing and/or treating cancers and
XX metastasis -
XX Disclosure; SEQ ID NO 41577; 3071pp + Sequence Listing; English.
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
XX amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
XX activity, and can be used in gene therapy and vaccine production. (I)
XX proteins and polynucleotides may be used in the prevention, diagnosis and
XX treatment of diseases associated with inappropriate (I) expression. For
XX example, they may be used to treat disorders associated with decreased
XX expression by rectifying mutations or deletions in a patient's genome
XX that affect the activity of (I) by expressing inactive proteins or to
XX supplement the patients own production of (I). Additionally, (I)
XX polynucleotides may be used to produce the secreted (I), by inserting
XX the nucleic acids into a host cell and culturing the cell to express the
XX protein. (I) proteins and polynucleotides may be used to prevent,
XX diagnose and treat immune/hematopoietic-related diseases, especially
XX cancers and cancer metastases of haematopoietic-derived cells. AAK64703
XX to AAK87694 represent human immune/hematopoietic antigen genomic
XX sequences from the present invention. AAK54942 to AAK54950 and AAK82169
XX represent sequences used in the exemplification of the present invention.
XX Sequence 20835 BP; 4759 A; 5358 C; 5249 G; 5469 T; 0 other;
Query Match 45.0%; Score 23.4; DB 22; Length 20835;
Best Local Similarity 67.3%; Pred. No. 75;
Matches 33; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
QY 1 aataaagattccattggataagaacgaatctgttacttgcgtgcag 49
Db 19017 AACATAATTCATTAGTTAAGAAATTTTCAGTTCTGCTGTGCAG 18969
RESULT 13
AAK6766/c
ID AAK86766 standard; DNA; 25806 BP.
XX AAK86766;
AC AAK86766;
XX 07-NOV-2001 (first entry)
XX Human immune/hematopoietic antigen genomic sequence SEQ ID NO:41578.
KW Human; immune; haematopoietic; immune/hematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX Homo sapiens.
XX WO200157182-A2.
XX 09-AUG-2001.
XX 17-JAN-2001; 2001WO-US01354.
XX 31-JAN-2000; 2000US-0179065.
```


FT	percentage identity to glycerol-3-phosphate dehydrogenase (GUT2) from <i>S. cerevisiae</i> "
FT	49377..49643
FT	/*tag= o
FT	/label= MG041
FT	/note= "The encoded protein shows 48.86 percentage identity to phosphohistidinoprotein-hexose phosphotransferase (ptsH) from <i>Mycoplasma capricolum</i> "
FT	50060..51520
FT	/*tag= p
FT	/label= MG042
FT	/note= "Previously identified as MORF-19832 and MORF-20108, the encoded protein shows 41.92 percentage identity to spermidine/putrescine transport ATP-binding protein (potA) from <i>E. coli</i> "
FT	51525..52382
FT	/*tag= q
FT	/label= MG043
FT	/note= "Previously identified as MORF-20110, the encoded protein shows 26.51 percentage identity to spermidine/putrescine transport system permease protein (potB) from <i>E. coli</i> "
FT	52366..53220
FT	/*tag= r
FT	/label= MG044
FT	/note= "Previously identified as MORF-20111, the encoded protein shows 29.45 percentage identity to spermidine/putrescine transport system permease protein (potC) from <i>E. coli</i> "
FT	54658..55605
FT	/*tag= s
FT	/label= MG046
FT	/note= "Previously identified as MORF-20112, the encoded protein shows 36.60 percentage identity to siologlycoprotease (gcp) from <i>Pasteurella haemolytica</i> complement (56970..58310)
FT	/*tag= t
FT	/label= MG048
FT	/note= "Previously identified as MORF-19834, MORF-20114 and MORF-20115, the encoded protein shows 43.02 percentage identity to signal recognition particle protein (ffh) from <i>B. subtilis</i> "
FT	58117..59079
FT	/*tag= u
FT	/label= MG049
FT	/note= "Previously identified as MORF-20114 and MORF-20115, the encoded protein shows 44.78 percentage identity to purine-nucleoside phosphorylase (deob) from <i>E. coli</i> "
FT	59083..59754
FT	/*tag= v
FT	/label= MG050
FT	/note= "Previously identified as MORF-20117, the encoded protein shows 83.03 percentage identity to deoxyribose-phosphate aldolase (deoc) from <i>Mycoplasma pneumoniae</i> "
FT	complement (64898..65731)
FT	/*tag= w
FT	/label= MG056
FT	/note= "Previously identified as MORF-20122, the encoded protein shows 30.25 percent identity to the protein disclosed in GB:D26185.99 from <i>B. subtilis</i> "
FT	complement (65713..66249)
FT	/*tag= x
FT	/label= MG057
FT	/note= "Previously identified as MORF-20123, the encoded protein shows 38.90 percentage identity to the protein disclosed in

```

FT          CDS          GB:D26185_104 from B. subtilis"
FT      81047..82597
FT      /*tag= y
FT      /label= MG067
FT      /note= "Previously identified as MORF-19845, the
FT      encoded protein shows 28.84 percentage
FT      identity to glutamic acid specific protease
FT      (SPase) from Staphylococcus aureus"
FT
FT      CDS          91065..91919
FT      /*tag= z
FT      /label= MG070
FT      /note= "Previously identified as MORF-20136, the
FT      encoded protein shows 34.8 percentage
FT      identity to ribosomal protein S2 (rps2)
FT      from Spirulina plantensis"
FT
FT
FT      Query Match          45.0%; Score 23.4; DB 18; Length 580073;
FT      Best Local Similarity 67.3%; Pred. NO. 1.3e+02;
FT      Matches 33; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
FT
FT
FT      QY      4      aaaaattcatttgataagaacgaattctgttactttgctgcagtc 52
FT      |||| | ||||| || | || | || |||| | | |||
FT      Db 438256 aaaaatacaccattgtattgctagcaattttcttgattccgtgc 438304
FT
FT
FT      RESULT 15
FT      AAH14861
FT      ID AAH14861 standard; cDNA; 2756 BP.
FT      XX
FT      AC AAH14861;
FT      XX
FT      DT 26-JUN-2001 (first entry)
FT      XX
FT      DE Human cDNA sequence SEQ ID NO:12702.
FT      XX
FT      KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
FT      OS Homo sapiens.
FT      XX
FT      PN EP1074617-A2.
FT      XX
FT      PD 07-FEB-2001.
FT      XX
FT      PF 28-JUL-2000; 2000EP-0116126.
FT      XX
FT      PR 29-JUL-1999; 99JP-0248036.
FT      PR 27-AUG-1999; 99JP-0300253.
FT      PR 11-JAN-2000; 2000JP-0118776.
FT      PR 02-MAY-2000; 2000JP-0183767.
FT      PR 09-JUN-2000; 2000JP-0241899.
FT      XX
FT      PA (HELI-) HELIX RES INST.
FT      XX
FT      PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
FT      PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
FT      XX
FT      DR WPI; 2001-318749/34.
FT      XX
FT      PR Primer sets for synthesizing polynucleotides, particularly the 5602
FT      full-length cDNAs defined in the specification, and for the detection
FT      and/or diagnosis of the abnormality of the proteins encoded by the
FT      full-length cDNAs -
FT      XX
FT      PS Claim 8; SEQ ID 12702; 2537pp + CD ROM; English.
FT      XX
FT      CC The present invention describes primer sets for synthesising 5602
FT      full-length cDNAs defined in the specification. Where a primer set
FT      comprises: (a) an oligo-dT primer and an oligonucleotide complementary
FT      to the complementary strand of a polynucleotide which comprises one of
FT      the 5602 nucleotide sequences defined in the specification, where the
FT      oligonucleotide comprises at least 15 nucleotides; or (b) a combination
FT      of an oligonucleotide comprising a sequence complementary to the
FT      complementary strand of a polynucleotide which comprises a 5'-end

```

CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesising polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.

XX
 SQ Sequence 2756 BP; 911 A; 438 C; 542 G; 865 T; 0 other;

Query Match 44.6%; Score 23.2; DB 22; Length 2756;
 Best Local Similarity 77.8%; Pred. No. 61;
 Matches 28; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 11 tccatttgatgaacgaatctgttactttgcttg 46
 ||||| ||||| ||||| |||||
 Db 1747 tccatttgatcagacacgtatgttccattgcttg 1782

Search completed: June 19, 2002, 17:15:17
 Job time: 18975 sec


```

; NAME/KEY: intron
; LOCATION: 912..1044
; FEATURE:
; NAME/KEY: intron
; LOCATION: 1155..3104
; FEATURE:
; NAME/KEY: intron
; LOCATION: 3376..4225
; FEATURE:
; NAME/KEY: intron
; LOCATION: 4329..4431
; FEATURE:
; NAME/KEY: intron
; LOCATION: 4778..4890
; FEATURE:
; NAME/KEY: intron
; LOCATION: 5254..5335
; FEATURE:
; NAME/KEY: intron
; LOCATION: 5415..5530
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(774..911, 1045..1154, 3105..3375, 4226..4328,
; LOCATION: 4432..4777, 4891..5253, 5336..5414, 5531..5743)
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 774..839
; FEATURE:
; NAME/KEY: polyA_site
; LOCATION: 6095..6100
; OTHER INFORMATION: /label= putative
;
US-08-687-580B-6

Query Match 44.6%; Score 23.2; DB 4; Length 6305;
Best Local Similarity 70.5%; Pred. No. 7.6; Indels 0;
Matches 31; Conservative 0; Mismatches 13;

Qy 2 ataaagattccatttgataagaacgaatctgttacttgcgtt 45
   | | | | | | | | | | | | | | | | | | | | | |
Db 1936 AAAAAGAGCTAGTTCGATATAGCGAGCTTTCGTTTTCTT 1979

RESULT 3
US-09-387-212-2
; Sequence 2, Application US/09387212A
; Patent No. 6309849
; GENERAL INFORMATION:
; APPLICANT: ROBISON, KEITH E.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN KINASE
; TITLE OF INVENTION: PHOSPHATASE HOMOLOGUES AND USES THEREFOR
; FILE REFERENCE: MNI-090
; CURRENT APPLICATION NUMBER: US/09/387,212A
; CURRENT FILING DATE: 1999-08-31
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 2
; LENGTH: 403
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: All occurrences of n indicate any nucleotide

US-09-387-212-2

Query Match 44.2%; Score 23; DB 4; Length 403;
Best Local Similarity 66.7%; Pred. No. 5.5;
Matches 33; Conservative 0; Mismatches 16; Indels 0;

Qy 3 taaagattccatttgataagaacgaatctgttactttgttcgact 50
   | | | | | | | | | | | | | | | | | | | | | |
Db 68 taaanaaacatttatgaaaaaagaacctttttcttaaatcagt 115

```

```
Query Match      43.8%; Score 22.8; DB 1; Length 8040;
Best Local Similarity 71.4%; Pred. NO. 11;
Matches 30; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
```

Qy 6 aagattccatttggtataagaacgaatctgttactttgctgc 47
 ||||| ||| |||| ||||| || |||| ||||| |||
 Db 6822 AAGATACCAGTTCGGGAACGAAGAGTTCGTTACATTGCCTGC 6863

RESULT 6
US-09-100-804-1
; Sequence 1, Application US/09100804
; Patent No. 6066472
; GENERAL INFORMATION:

US-09-100-804-1
; Sequence 1, Application US/09100804
; Patent No. 6066472
; GENERAL INFORMATION:
; APPLICANT: GONEZ, LEONEL JORGE
; APPLICANT: SARAS, JAN
; APPLICANT: CLAESSON-WELSH, LENA
; APPLICANT: HELDIN, CARL-HENRIK
; TITLE OF INVENTION: PRIMARY STRUCTURE AND FUNCTIONAL
; TITLE OF INVENTION: EXPRESSION OF NUCLEOTIDE SEQUENCES FOR NOVEL PROTEIN
; TITLE OF INVENTION: TYROSINE PHOSPHATASES
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

```
;
; REFERENCE/DOCKET NUMBER: L0461/7003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8040 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna to mRNA
; HYPOTHETICAL: NO.
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: HOMO SAPIENS
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 78..7475
; PCT-US94-09943-1
;
; Query Match 43.8%; Score 22.8; DB 3; Length 8040;
; Best Local Similarity 71.4%; Pred. No. 11;
; Matches 30; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
;
; QY 6 aagattccattgataagaacgaatctgttacttctgttc 47
; ||||| ||||| ||||| ||||| ||||| ||||| |||||
; DB 6822 AAGATACCAAGTTGGGAAAGAGAGTTCGTTTACATTCCCTGC 6863
;
; RESULT 7
; PCT-US94-09943-1
; Sequence 1, Application PC/TUS9409943
; GENERAL INFORMATION:
; APPLICANT:
; APPLICANT:
; APPLICANT:
; APPLICANT:
; TITLE OF INVENTION: PRIMARY STRUCTURE AND FUNCTIONAL
; TITLE OF INVENTION: EXPRESSION OF NUCLEOTIDE SEQUENCES FOR NOVEL PROTEIN
; TITLE OF INVENTION: TYROSINE PHOSPHATASES
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; STREET: 600 ATLANTIC AVENUE
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: USA
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/09943
; FILING DATE: 01-SEP-1994
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/115,573
; FILING DATE: 01-SEP-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: TWOMEY, MICHAEL J.
; REGISTRATION NUMBER: P-38,349
; REFERENCE/DOCKET NUMBER: L0461/7000WO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/720-3500
; TELEFAX: 617/720-2441
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8043 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna to mRNA
; HYPOTHETICAL: NO.
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: HOMO SAPIENS
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 78..7478
; PCT-US94-09943-1
;
; Query Match 43.8%; Score 22.8; DB 4; Length 8119;
; Best Local Similarity 71.4%; Pred. No. 11;
; Matches 30; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
;
; QY 6 aagattccattgataagaacgaatctgttacttctgttc 47
; ||||| ||||| ||||| ||||| ||||| ||||| |||||
; DB 6868 aagataccagttgggaagaagagttcgtttacattgctgc 6909
;
; RESULT 8
; US-09-290-640-45
; Sequence 45, Application US/09290640
; Patent No. 6204055
; GENERAL INFORMATION:
; APPLICANT: Dean, Nicholas M.
; APPLICANT: Marcussen, Eric G.
; TITLE OF INVENTION: Antisense Compound Modulation of Fas Mediated Signaling
; FILE REFERENCE: ISPH-0351
; CURRENT APPLICATION NUMBER: US/09/290,640
; CURRENT FILING DATE: 1999-04-12
; NUMBER OF SEQ ID NOS: 85
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 45
; LENGTH: 8119
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (64)..(7521)
; PUBLICATION INFORMATION:
; JOURNAL: FEBS Lett.
; VOLUME: 337
; ISSUE: 2
; PAGES: 200-206
; DATE: 1994-01-10
; DATABASE ACCESSION NUMBER: D21209/Genbank
; DATABASE ENTRY DATE: 1999-02-05
; US-09-290-640-45
;
; Query Match 43.8%; Score 22.8; DB 4; Length 8119;
; Best Local Similarity 71.4%; Pred. No. 11;
; Matches 30; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
;
; QY 6 aagattccattgataagaacgaatctgttacttctgttc 47
; ||||| ||||| ||||| ||||| ||||| ||||| |||||
; DB 6868 aagataccagttgggaagaagagttcgtttacattgctgc 6909
;
; RESULT 9
; US-09-341-587-7/c
; Sequence 7, Application US/09341587
; Patent No. 6346606
; GENERAL INFORMATION:
; APPLICANT: Mollenhauer, Jan
; TITLE OF INVENTION: Protein Containing an SRCR Domain
; FILE REFERENCE: 4121-108
; CURRENT APPLICATION NUMBER: US/09/341,587
; CURRENT FILING DATE: 1999-08-31
; EARLIER APPLICATION NUMBER: PCT/DE98/00096
```

100

```

Query Match          41.5%; Score 21.6; DB 2; Length 3959;
Best Local Similarity 68.2%; Pred. NO. 27;
Matches 30; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Qy      8 gatttcattgataagaacgaatctgttactttgctgcagtg 51
        ||| ||||| ||| ||||| | ||||| ||
Db    3318 GAATACACTAGATATAGGCACTGTTCCTCCTTGCACGTG 3275

RESULT 13
US-09-268-992-7
; Sequence 7, Application US/09268992
; Patent No. 6342351
; GENERAL INFORMATION:
; APPLICANT: Chen, H.
; APPLICANT: Feilmer, N.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CHROMOSOME-18p RELATED DISORDERS
; FILE REFERENCE: 7853-138
; CURRENT APPLICATION NUMBER: US/09/268,992
; CURRENT FILING DATE: 1999-03-16
; EARLIER APPLICATION NUMBER: 09/236,134
; EARLIER FILING DATE: 1999-01-22
; EARLIER APPLICATION NUMBER: 60/106,056
; EARLIER FILING DATE: 1998-10-28
; EARLIER APPLICATION NUMBER: 60/088,312
; EARLIER FILING DATE: 1998-06-05
; EARLIER APPLICATION NUMBER: 60/078,044
; EARLIER FILING DATE: 1998-03-16
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 72604
; TYPE: DNA
; ORIGIN: Homo sapiens

```

```

; FILING DATE: 06-FEB-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/494,577
; FILING DATE: 22-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 05433/032001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-542-5070
; TELEFAX: 617-542-8906
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2738 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
US-08-795-868-17

```

```

Query Match          41.2%; Score 21.4; DB 2; Length 2738;
Best Local Similarity 66.0%; Pred. No. 30;
Matches 31; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Qy  2 ataaaagattccattggataagaacgaatctgttactttgcttgca 48
    ||| |||| | |||| | |||| | |||| | |||| | |||| |
Db  489 ATATAAGAGCATTTTGAAAAAAACTAATGTTTGAAATCGCTGGCA 535

```

Search completed: June 19, 2002, 16:33:08
Job time: 16446 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2002, 14:04:31 ; Search time 7489.97 seconds
(without alignments)
93.704 Million cell updates/sec

Title: US-09-462-955B-1_COPY_734_785

Perfect score: 52

Sequence: 1 aataaaagattccattgga.....tggtactttgttcgactgc 52

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : EST:*

- 1: em_estba:*
- 2: em_esthum:*
- 3: em_estin:*
- 4: em_estmu:*
- 5: em_estov:*
- 6: em_estpl:*
- 7: em_estro:*
- 8: em_hic:*
- 9: gb_est1:*
- 10: gb_est2:*
- 11: gb_hic:*
- 12: gb_gss:*
- 13: em_gss_hum:*
- 14: em_gss_inv:*
- 15: em_gss_pln:*
- 16: em_gss_vrt:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	26.4	50.8	257	12	AQ106136
c 2	26.2	50.4	999	10	BG749262
c 3	26	50.0	864	12	A2547272
c 4	25.8	49.6	418	9	A1065614
5	25.8	49.6	487	12	B43802
6	25.8	49.6	537	12	AQ694110
c 7	25.4	48.8	525	12	A2390369
8	25.4	48.8	576	10	BG654372
9	25.4	48.8	781	10	BE884333
10	25	48.1	433	10	BE678889
11	25	48.1	456	10	BE678999
c 12	25	48.1	487	12	AQ270104
13	25	48.1	501	10	BE678820
14	25	48.1	560	10	BE678247
c 15	25	48.1	625	12	AQ020281
16	25	48.1	811	9	AU122315
17	25	48.1	904	12	CNS049BD

c 18	24.8	47.7	577	12	AQ506286
c 19	24.8	47.7	648	12	AG160388
c 20	24.6	47.3	516	10	BM198468
21	24.6	47.3	713	12	AG089691
c 22	24.4	46.9	321	9	AA221297
23	24.4	46.9	365	10	BF547464
24	24.4	46.9	367	10	BF016895
25	24.4	46.9	385	10	BE688287
c 26	24.4	46.9	390	9	A1021393
27	24.4	46.9	412	12	AZ010087
c 28	24.4	46.9	421	9	BB675436
c 29	24.4	46.9	650	9	AV344344
c 30	24.4	46.9	708	10	BT157582
c 31	24.4	46.9	747	12	CNS03244
c 32	24.2	46.5	317	10	BF747486
33	24.2	46.5	375	9	AA228092
34	24.2	46.5	394	9	AW037099
35	24.2	46.5	398	12	AQ135393
c 36	24.2	46.5	418	10	BF728680
c 37	24.2	46.5	507	9	AW566513
c 38	24.2	46.5	517	9	AV834321
c 39	24.2	46.5	565	10	BM382402
c 40	24.2	46.5	626	10	BM380636
c 41	24.2	46.5	628	10	BG842500
42	24.2	46.5	650	9	AW303369
c 43	24.2	46.5	711	10	BF630600
c 44	24.2	46.5	1101	12	CNS00KH0
c 45	24	46.2	312	10	BF062027

ALIGNMENTS

RESULT 1

AQ106136
LOCUS
DEFINITION HS_3055_A2_B07_MR CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=3055 Col=14 Row=C, DNA sequence.
ACCESSION AQ106136.1 GI:3481492
VERSION GSS.
KEYWORDS
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 257)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 3055 row: C column: 14
Class: BAC ends
High quality sequence stop: 257.
Location/Qualifiers
1. 257
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=3055 Col=14 Row=C"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH10B"

```

BASE COUNT      93 a      30 c      43 g      91 t
ORIGIN

Query Match      50.8%; Score 26.4; DB 12; Length 257;
Best Local Similarity 75.0%; Pred. No. 43;
Matches 33; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 4 aaagattccatttgataagaacgaactgttacttgcgttc 47
||||| ||||| ||| ||| ||||| ||||| ||||| |||||
Db 48 AAAATAATACATTGGGTAGGTCTAATCTGTTACTGTTATTGC 91

RESULT 2
BG749262/c
LOCUS
DEFINITION 602708092f1 NIH_MGC_43 Homo sapiens cDNA clone IMAGE:4844615 5',
mRNA sequence.
ACCESSION BG749262
VERSION BG749262.1 GI:14059915
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 999)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1682 row: h column: 24
High quality sequence stop: 810.
FEATURES
Location/Qualifiers
1..999
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4844615"
/clone_lib="NIH_MGC_43"
/tissue_type="normal pigmented retinal epithelium"
/lab_host="DH10B (phage-resistant)"
/note="Organ: eye; Vector: pORF7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library. |"
BASE COUNT      213 a      267 c      311 g      208 t
ORIGIN

Query Match      50.4%; Score 26.2; DB 10; Length 999;
Best Local Similarity 79.5%; Pred. No. 58;
Matches 31; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 8 gattccatttgataagaacgaactgttacttgcgttc 46
||||| ||||| ||| ||| ||||| ||||| ||||| |||||
Db 589 GATTCCATCTGTTATGAGGAGTGTACCACCTTGCCTTC 551

RESULT 3
AZ547272/c
LOCUS
DEFINITION ENT0220TR Entamoeba histolytica Sheared DNA Entamoeba histolytica

genomic, DNA sequence.
ACCESSION AZ547272
VERSION AZ547272.1 GI:11169813
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica.
Eukaryota; Entamoebidae; Entamoeba.
REFERENCE 1 (bases 1 to 864)
AUTHORS Loftus,B., Van Aken,S. and Fraser,C.
TITLE Determination of clone end sequences from Entamoeba histolytica
JOURNAL HMI:IMSS sheared DNA library
COMMENT Unpublished (2000)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjoftus@tigr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
DNA library
Seq primer: M13-Reverse
Class: shotgun
High quality sequence start: 16
High quality sequence stop: 811.
High quality sequence stop: 811.
Location/Qualifiers
1..864
/organism="Entamoeba histolytica"
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: pHOS1; Site_1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + i method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barell, Oxford University Press, 1999)."
BASE COUNT      352 a      103 c      90 g      319 t
ORIGIN

Query Match      50.0%; Score 26; DB 12; Length 864;
Best Local Similarity 70.0%; Pred. No. 67;
Matches 35; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 1 aataaagattccatttgataagaacgaactgttacttgcgttc 50
||||| ||||| ||||| ||| ||||| ||||| ||||| |||||
Db 604 AATAAAGATATTATTGTTGTTAAATTTGTAATTTGGTAATTTGTTGTTAT 555

RESULT 4
AI065614/c
LOCUS
DEFINITION ag90e12.x1 maize inflorescence immature ear library Zea mays cDNA
clone ag90e12 3', mRNA sequence.
ACCESSION AI065614
VERSION AI065614.1 GI:3341021
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 416)
AUTHORS Schut, K., de la Bastide, M., Gnoj, L., Habermann, K., Huang, E. N.,
Parnell, L. D., Dedhia, N., Martienssen, R. and McCombie, W. R.

```

<p>TITLE Expressed sequence tags from Z. mays</p> <p>JOURNAL Unpublished (1998)</p> <p>COMMENT Contact: W. Richard McCombie Lita Annenberg Hazen Genome Sequencing Center Cold Spring Harbor Laboratory PO Box 100, Cold Spring Harbor, NY 11724, USA Tel: 516 367 8884 Fax: 516 367 8874 Email: mcombie@cshl.org Plate: ag90 row: e column: 12 Seq primer: M13 forward universal -21 High quality sequence stop: 418.</p>	<p>BASE COUNT 121 a 98 c 78 g 121 t</p> <p>ORIGIN</p>	<p>BASE COUNT 153 a 110 c 112 g 161 t 1 others</p> <p>ORIGIN</p>
<p>FEATURES source</p>	<p>FEATURES source</p>	<p>FEATURES source</p>
<p>REFERENCE 1 (bases 1 to 487)</p> <p>AUTHORS Mahairas,G.G., Zackrone,K.D., Smith,T., Tipton,S., Schmidt,S., Traicoff,R., Abajian,C., Blanchard,A., West,A. and Hood,L.E.</p> <p>TITLE Construction of a Characterized Clone Resource for Genomic Sequencing: Generation and Preliminary Analysis of 20,000 Sequence Tagged Connectors</p> <p>JOURNAL Unpublished (1997)</p> <p>COMMENT Contact: Mahairas GG, Zackrone KD, Hood L University of Washington Seattle, WA 98195, USA Tel: (206) 616-8744 Fax: (206) 685-7301 Email: kzackron@u.washington.edu Sequence Tagged Connector Plate: CT 780 row: G column: 12 Class: BAC ends High quality sequence stop: 487.</p>	<p>REFERENCE 1 (bases 1 to 537)</p> <p>AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.</p> <p>TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome</p> <p>JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)</p> <p>MEDLINE 99380589</p> <p>COMMENT Contact: Mahairas GG, Wallace JC, Hood L University of Washington 401 Queen Anne Avenue North, Seattle, WA 98109, USA Tel: (206) 616-3618 Fax: (206) 616-3887 Email: jwallace@u.washington.edu Clones may be purchased from Research Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu Plate: 2100 row: O column: 12 Seq primer: T7 Class: BAC ends High quality sequence stop: 537.</p>	<p>REFERENCE 1 (bases 1 to 537)</p> <p>AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.</p> <p>TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome</p> <p>JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)</p> <p>MEDLINE 99380589</p> <p>COMMENT Contact: Mahairas GG, Wallace JC, Hood L University of Washington 401 Queen Anne Avenue North, Seattle, WA 98109, USA Tel: (206) 616-3618 Fax: (206) 616-3887 Email: jwallace@u.washington.edu Clones may be purchased from Research Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu Plate: 2100 row: O column: 12 Seq primer: T7 Class: BAC ends High quality sequence stop: 537.</p>
<p>REFERENCE 1 (bases 1 to 537)</p> <p>AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.</p> <p>TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome</p> <p>JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)</p> <p>MEDLINE 99380589</p> <p>COMMENT Contact: Mahairas GG, Wallace JC, Hood L University of Washington 401 Queen Anne Avenue North, Seattle, WA 98109, USA Tel: (206) 616-3618 Fax: (206) 616-3887 Email: jwallace@u.washington.edu Clones may be purchased from Research Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu Plate: 2100 row: O column: 12 Seq primer: T7 Class: BAC ends High quality sequence stop: 537.</p>	<p>REFERENCE 1 (bases 1 to 537)</p> <p>AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.</p> <p>TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome</p> <p>JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)</p> <p>MEDLINE 99380589</p> <p>COMMENT Contact: Mahairas GG, Wallace JC, Hood L University of Washington 401 Queen Anne Avenue North, Seattle, WA 98109, USA Tel: (206) 616-3618 Fax: (206) 616-3887 Email: jwallace@u.washington.edu Clones may be purchased from Research Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu Plate: 2100 row: O column: 12 Seq primer: T7 Class: BAC ends High quality sequence stop: 537.</p>	<p>REFERENCE 1 (bases 1 to 537)</p> <p>AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.</p> <p>TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome</p> <p>JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)</p> <p>MEDLINE 99380589</p> <p>COMMENT Contact: Mahairas GG, Wallace JC, Hood L University of Washington 401 Queen Anne Avenue North, Seattle, WA 98109, USA Tel: (206) 616-3618 Fax: (206) 616-3887 Email: jwallace@u.washington.edu Clones may be purchased from Research Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu Plate: 2100 row: O column: 12 Seq primer: T7 Class: BAC ends High quality sequence stop: 537.</p>
<p>REFERENCE 1 (bases 1 to 537)</p> <p>AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.</p> <p>TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome</p> <p>JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)</p> <p>MEDLINE 99380589</p> <p>COMMENT Contact: Mahairas GG, Wallace JC, Hood L University of Washington 401 Queen Anne Avenue North, Seattle, WA 98109, USA Tel: (206) 616-3618 Fax: (206) 616-3887 Email: jwallace@u.washington.edu Clones may be purchased from Research Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu Plate: 2100 row: O column: 12 Seq primer: T7 Class: BAC ends High quality sequence stop: 537.</p>	<p>REFERENCE 1 (bases 1 to 537)</p> <p>AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.</p> <p>TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome</p> <p>JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)</p> <p>MEDLINE 99380589</p> <p>COMMENT Contact: Mahairas GG, Wallace JC, Hood L University of Washington 401 Queen Anne Avenue North, Seattle, WA 98109, USA Tel: (206) 616-3618 Fax: (206) 616-3887 Email: jwallace@u.washington.edu Clones may be purchased from Research Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu Plate: 2100 row: O column: 12 Seq primer: T7 Class: BAC ends High quality sequence stop: 537.</p>	<p>REFERENCE 1 (bases 1 to 537)</p> <p>AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.</p> <p>TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome</p> <p>JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)</p> <p>MEDLINE 99380589</p> <p>COMMENT Contact: Mahairas GG, Wallace JC, Hood L University of Washington 401 Queen Anne Avenue North, Seattle, WA 98109, USA Tel: (206) 616-3618 Fax: (206) 616-3887 Email: jwallace@u.washington.edu Clones may be purchased from Research Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu Plate: 2100 row: O column: 12 Seq primer: T7 Class: BAC ends High quality sequence stop: 537.</p>
<p>REFERENCE 1 (bases 1 to 537)</p> <p>AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.</p> <p>TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome</p> <p>JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)</p> <p>MEDLINE 99380589</p> <p>COMMENT Contact: Mahairas GG, Wallace JC, Hood L University of Washington 401 Queen Anne Avenue North, Seattle, WA 98109, USA Tel: (206) 616-3618 Fax: (206) 616-3887 Email: jwallace@u.washington.edu Clones may be purchased from Research Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu Plate: 2100 row: O column: 12 Seq primer: T7 Class: BAC ends High quality sequence stop: 537.</p>	<p>REFERENCE 1 (bases 1 to 537)</p> <p>AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.</p> <p>TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome</p> <p>JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-974</p>	

```

RESULT      7
AZ390369/c
LOCUS
DEFINITION  AZ390369          525 bp        DNA            linear       GSS 03-OCT-2000
              MW015M10R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC1M015M10 R, DNA sequence.

ACCESSION   AZ390369
VERSION     AZ390369.1  GI:10505412
KEYWORDS    GSS.
SOURCE      house mouse.
ORGANISM    Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 525)
AUTHORS     Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
             Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
             ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.,
             and Wright,D., Weiss,R.
TITLE       Mouse whole genome scaffolding with paired end reads from 10kb
             plasmid inserts
JOURNAL     Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
             University of Utah Genome Center
             University of Utah
             Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
             84112, USA
             Tel.: 801 585 5606
             Fax: 801 585 7177
             Email: ddunn@genetics.utah.edu
             Insert Length: 10000 Std Error: 0.00
             Plate: 0151 row: M column: 10
             Seq primer: CACACAGGAACACGCTATGACC
             Class: plasmid ends
             High quality sequence stop: 525.
FEATURES             Location/Qualifiers
                     1..525
                        /organism="Mus musculus"
                        /strain="C57BL/6J"
                        /db_xref="taxon:10090"
                        /clone="UUGC1M015M10"
                        /clone_lib="Mouse 10kb plasmid UUGC1M library"
                        /sex="Male"
                        /lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
                        /notes="Vector: pWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT      177 a   100 c   88 g   160 t
ORIGIN
Query Match      48.8%; Score 25.4; DB 12; Length 525;
Best Local Similarity 68.6%; Pred. No. 1e+02;
Matches 35; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Qy 2 ataagaattccatttggaagaacaagaaatcttacttcgttcgcagtcg 52
|||||| | | ||| |||| | | | | | | | | | | | | | | | | | | | |

```


!

PCR) to Cot-omega of 11. After removal of hybrids and excess drier by streptavidin sepharose chromatography, the ss-phagemids were made double stranded and electroporated into Top-10 F'. Original library constructed by Bruce Blumberg (Blumberg et al., 1991 Science 253, 194-196; Hawlet et al., 1995, Genes Dev. 9, 2923-2935). Note: This is a Xenopus Gene Collection (XGC library).

Query Match 48.1%; Score 25; DB 10; Length 560;
Best Local Similarity 69.4%; Pred. No. 1.4e+02;
Matches 34; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 4 aaagattccatttgataagaacgaatctgttacttcttgcagtcg 52
Db 208 AAAGGGTGTATTGGCAAGGAAAAACCAGATATTTTGGTTGCAGTGC 256

LOCUS	AQ020281	625 bp	DNA	linear	GSS 09-JUN-1998
DEFINITION	CIT-HSP-2307A3.TF CIT-HSP Homo sapiens genomic clone 2307A3, DNA sequence.				

SOURCE	ORGANISM	TITLE
Human.	Homo sapiens	Building (1998)
Eukaryota;	Metazoa;	Chordata;
Mammalia;	Eutheria;	Primates;
1 (bases 1 to 625)		Catarrhini; Hominoidea; Homo.
ADAMS, M.D.,	ROUNSELEY, S.D.,	ZHAO, S.,
GOLDEN, K.,	BERRY, K.,	GRANGER, D.,
SIMON, M. and	VENTER, J.C.	
Use of a random BAC End Sequence Database for Sequence-Ready Map Building (1998)		

Contact: Mark Adams
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: madams@igmc.org

email: mduadins@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13-21
Class: BAC ends.

```

/clones="2307A3"
/clone_lib="CIT-HSP"
/sex="Male"
/cell_type="Sperm"
/notes="Vector: pBelOAC11; Site_1: HindIII; Site_2:
HindIII"
BASE COUNT      225 a      103 g      192 t
ORIGIN

Query Match      48.1%; Score 25; DB 12; Length 625;
Best Local Similarity 59.4%; Pred. No. 1.4e+02;
Matches 34; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy      2 ataaagattccatttgataagaacgaatctgttacttgcgtcagt 50
      |||||

```

Db 262 AAAAATGGTTACATTTAAAAAGATCGATTCCCTTCCTTTTCTTACATT 214

Search completed: June 19, 2002, 14:04:35
Job time: 7533 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2002, 11:59:02 ; Search time 247.21 Seconds
(without alignments)
21.860 Million cell updates/sec

Title: US-09-462-955B-1_COPY_655_676

Perfect score:

Sequence: 1 ctgccaggccgaaggcctggg 22

Scoring table: IDENTITY NUC

IDENTITY_NOC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

```
Maximum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%

Post-processing: Minimum Match 0%
Maximum Match 100%

Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:**

```

database ;
Issued_Patents_NA:
1: /cqn2_6/ptodata/1/ina/5A COMB.seq.*

```

```
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
```

```
2: /cgn2_0/ptodata/1/ina/3B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
```

```
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
```

```
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
```

```
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query			ID	Description
		Match	Length	DB		
1	22	100.0	23	4	US-09-462-975-6	Sequence 6, Appli
2	22	100.0	1291	4	US-09-462-975-1	Sequence 1, Appli
c 3	17.8	80.9	1938	4	US-09-232-200-29	Sequence 29, Appli
c 4	17.8	80.9	1938	4	US-09-232-197-29	Sequence 29, Appli
c 5	17.8	80.9	1938	4	US-09-232-201-29	Sequence 29, Appli
c 6	17.8	80.9	1941	4	US-09-232-200-28	Sequence 28, Appli
c 7	17.8	80.9	1941	4	US-09-232-197-28	Sequence 28, Appli
c 8	17.8	80.9	1941	4	US-09-232-201-28	Sequence 28, Appli
c 9	17.8	80.9	2816	1	US-08-785-241-1	Sequence 1, Appli
c 10	17.8	80.9	3217	4	US-09-232-200-64	Sequence 64, Appli
c 11	17.8	80.9	3217	4	US-09-232-197-64	Sequence 64, Appli
c 12	17.8	80.9	3217	4	US-09-232-201-64	Sequence 64, Appli
c 13	17.8	80.9	3694	4	US-09-232-200-46	Sequence 46, Appli
c 14	17.8	80.9	3694	4	US-09-232-197-46	Sequence 46, Appli
c 15	17.8	80.9	3694	4	US-09-232-201-46	Sequence 46, Appli
c 16	17.8	80.9	3704	4	US-09-232-200-24	Sequence 24, Appli
c 17	17.8	80.9	3704	4	US-09-232-197-24	Sequence 24, Appli
c 18	17.8	80.9	3704	4	US-09-232-201-24	Sequence 24, Appli
c 19	17.2	78.2	1431	4	US-09-491-772-7	Sequence 7, Appli
c 20	17.2	78.2	3131	3	US-09-035-648-23	Sequence 23, Appli
c 21	17.2	78.2	3131	4	US-09-001-951-23	Sequence 23, Appli
c 22	17.2	78.2	8580	4	US-09-491-772-1	Sequence 1, Appli
c 23	16.4	74.5	4411529	4	US-09-103-840A-1	Sequence 1, Appli
c 24	16.2	73.6	472	2	US-08-811-949-40	Sequence 40, Appli
c 25	16.2	73.6	747	1	US-08-257-341-6	Sequence 6, Appli
c 26	16.2	73.6	779	1	US-08-133-804-3	Sequence 3, Appli
c 27	16.2	73.6	779	1	US-08-461-838-3	Sequence 3, Appli

ALIGNMENTS

```

RESULT      1
US-09-462-975-6
; Sequence 6, Application US/09462975
; Patent No. 6103345
; GENERAL INFORMATION:
; APPLICANT: Rohde, Wolfgang
; APPLICANT: Becker, Dieter
; APPLICANT: Randles, John W.
; APPLICANT: Hehn, Alain
; APPLICANT: Salamini, Francesco
; TITLE OF INVENTION: USE OF VIRUS DNA AS A PROMOTER
; FILE REFERENCE: 23232.0003U1
; CURRENT APPLICATION NUMBER: US/09/462,975
; CURRENT FILING DATE: 2000-05-17
; PRIOR APPLICATION NUMBER: PCT/EP98/04345
; PRIOR FILING DATE: 1998-07-13
; PRIOR APPLICATION NUMBER: 19730502.4
; PRIOR FILING DATE: 1997-07-16
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 23
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence
US-09-462-975-6

```

Sequence 3, Appl
Sequence 1, Appl
Sequence 1, Appl
Sequence 5, Appl
Sequence 60, Appl
Sequence 2, Appl
Sequence 3, Appl
Sequence 44, Appl
Sequence 46, Appl
Sequence 52, Appl
Sequence 58, Appl
Sequence 58, Appl
Sequence 4, Appl
Sequence 9, Appl
Sequence 64, Appl
Sequence 66, Appl
Sequence 1, Appl
Sequence 1, Appl

FILE REFERENCE: 23232.0003U1
CURRENT APPLICATION NUMBER: US/09/462,975
CURRENT FILING DATE: 2000-05-17
PRIOR APPLICATION NUMBER: PCT/EP98/04345
PRIOR FILING DATE: 1998-07-13
PRIOR APPLICATION NUMBER: 19730502.4
PRIOR FILING DATE: 1997-07-16
NUMBER OF SEQ ID NOS: 7
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 1291
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence;/note-synthetic
US-09-462-975-1

Query Match 100.0%; Score 22; DB 4; Length 1291;
Best Local Similarity 100.0%; Pred. No. 0.56; 0; Indels 0; Gaps 0;
Matches 22; Conservative 0; Mismatches 0;

QY 1 ctgccagggccgaaggcctggg 22
|||||
Db 1045 ctgccagggccgaaggcctggg 1066

RESULT 3
US-09-232-200-29/c
Sequence 29, Application US/09232200A
Patent No. 6288213
GENERAL INFORMATION:
APPLICANT: Stahl, Andreas
APPLICANT: Hirsch, David J.
APPLICANT: Lodish, Harvey F.
APPLICANT: Gimeno, Ruth E.
APPLICANT: Tartaglia, Louis A.
TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
FILE REFERENCE: WHI97-21p3MB
CURRENT APPLICATION NUMBER: US/09/232,200A
CURRENT FILING DATE: 1999-01-14
EARLIER APPLICATION NUMBER: 60/071,374
EARLIER FILING DATE: 1998-01-15
EARLIER APPLICATION NUMBER: 60/093,491
EARLIER FILING DATE: 1998-07-20
EARLIER APPLICATION NUMBER: 60/110,941
EARLIER FILING DATE: 1998-12-04
NUMBER OF SEQ ID NOS: 105
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 29
LENGTH: 1938
TYPE: DNA
ORGANISM: Mus musculus
US-09-232-200-29

Query Match 80.9%; Score 17.8; DB 4; Length 1938;
Best Local Similarity 90.5%; Pred. No. 30;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 tgcccagggccgaaggcctggg 22
|||||
Db 517 TGCCAGGCGAGAAGGCCAGGG 497

RESULT 4
US-09-232-197-29/c
Sequence 29, Application US/09232197A
Patent No. 6300096
GENERAL INFORMATION:
APPLICANT: Stahl, Andreas
APPLICANT: Hirsch, David J.

APPLICANT: Lodish, Harvey F.
APPLICANT: Gimeno, Ruth E.
APPLICANT: Tartaglia, Louis A.
TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
FILE REFERENCE: WHI97-21p3MA
CURRENT APPLICATION NUMBER: US/09/232,197A
CURRENT FILING DATE: 1999-01-14
EARLIER APPLICATION NUMBER: 60/071,374
EARLIER FILING DATE: 1998-01-15
EARLIER APPLICATION NUMBER: 60/093,491
EARLIER FILING DATE: 1998-07-20
EARLIER APPLICATION NUMBER: 60/110,941
EARLIER FILING DATE: 1998-12-04
NUMBER OF SEQ ID NOS: 105
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 29
LENGTH: 1938
TYPE: DNA
ORGANISM: Mus musculus
US-09-232-197-29

Query Match 80.9%; Score 17.8; DB 4; Length 1938;
Best Local Similarity 90.5%; Pred. No. 30;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 tgcccagggccgaaggcctggg 22
|||||
Db 517 TGCCAGGCGAGAAGGCCAGGG 497

RESULT 5
US-09-232-201-29/c
Sequence 29, Application US/09232201A
Patent No. 6348321
GENERAL INFORMATION:
APPLICANT: Stahl, Andreas
APPLICANT: Hirsch, David J.
APPLICANT: Lodish, Harvey F.
APPLICANT: Gimeno, Ruth E.
APPLICANT: Tartaglia, Louis A.
TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
FILE REFERENCE: WHI97-21p3MC
CURRENT APPLICATION NUMBER: US/09/232,201A
CURRENT FILING DATE: 1999-01-14
EARLIER APPLICATION NUMBER: 60/071,374
EARLIER FILING DATE: 1998-01-15
EARLIER APPLICATION NUMBER: 60/093,491
EARLIER FILING DATE: 1998-07-20
EARLIER APPLICATION NUMBER: 60/110,941
EARLIER FILING DATE: 1998-12-04
NUMBER OF SEQ ID NOS: 105
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 29
LENGTH: 1938
TYPE: DNA
ORGANISM: Mus musculus
US-09-232-201-29

Query Match 80.9%; Score 17.8; DB 4; Length 1938;
Best Local Similarity 90.5%; Pred. No. 30;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 tgcccagggccgaaggcctggg 22
|||||
Db 517 TGCCAGGCGAGAAGGCCAGGG 497

RESULT 6
US-09-232-200-28/c
Sequence 28, Application US/09232200A
Patent No. 6288213

US-09-232-201-28/c
; GENERAL INFORMATION:
; APPLICANT: Stahl, Andreas
; APPLICANT: Hirsch, David J.
; APPLICANT: Lodish, Harvey F.
; APPLICANT: Gimeno, Ruth E.
; APPLICANT: Tartaglia, Louis A.
; TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
; FILE REFERENCE: WHI97-21p3MB
; CURRENT APPLICATION NUMBER: US/09/232,200A
; CURRENT FILING DATE: 1999-01-14
; EARLIER APPLICATION NUMBER: 60/071,374
; EARLIER FILING DATE: 1998-01-15
; EARLIER APPLICATION NUMBER: 60/093,491
; EARLIER FILING DATE: 1998-07-20
; EARLIER APPLICATION NUMBER: 60/110,941
; EARLIER FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 28
; LENGTH: 1941
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-232-200-28

Query Match 80.9%; Score 17.8; DB 4; Length 1941;
Best Local Similarity 90.5%; Pred. No. 30;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 tgcccaggccgaagcctggg 22
||||||| ||||| |||
Db 517 TGCCAGCAGAGGCCAGGG 497

RESULT 7

US-09-232-197-28/c
; Sequence 28, Application US/09232197A
; Patent No. 6300096
; GENERAL INFORMATION:
; APPLICANT: Stahl, Andreas
; APPLICANT: Hirsch, David J.
; APPLICANT: Lodish, Harvey F.
; APPLICANT: Gimeno, Ruth E.
; APPLICANT: Tartaglia, Louis A.
; TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
; FILE REFERENCE: WHI97-21p3MA
; CURRENT APPLICATION NUMBER: US/09/232,197A
; CURRENT FILING DATE: 1999-01-14
; EARLIER APPLICATION NUMBER: 60/071,374
; EARLIER FILING DATE: 1998-01-15
; EARLIER APPLICATION NUMBER: 60/093,491
; EARLIER FILING DATE: 1998-07-20
; EARLIER APPLICATION NUMBER: 60/110,941
; EARLIER FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 28
; LENGTH: 1941
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-232-197-28

Query Match 80.9%; Score 17.8; DB 4; Length 1941;
Best Local Similarity 90.5%; Pred. No. 30;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 tgcccaggccgaagcctggg 22
||||||| ||||| |||
Db 517 TGCCAGCAGAGGCCAGGG 497

RESULT 8

US-09-232-201-28/c
; Sequence 28, Application US/09232201A
; Patent No. 6348321
; GENERAL INFORMATION:
; APPLICANT: Stahl, Andreas
; APPLICANT: Hirsch, David J.
; APPLICANT: Lodish, Harvey F.
; APPLICANT: Gimeno, Ruth E.
; APPLICANT: Tartaglia, Louis A.
; TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
; FILE REFERENCE: WHI97-21p3MC
; CURRENT APPLICATION NUMBER: US/09/232,201A
; CURRENT FILING DATE: 1999-01-14
; EARLIER APPLICATION NUMBER: 60/071,374
; EARLIER FILING DATE: 1998-01-15
; EARLIER APPLICATION NUMBER: 60/093,491
; EARLIER FILING DATE: 1998-07-20
; EARLIER APPLICATION NUMBER: 60/110,941
; EARLIER FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 28
; LENGTH: 1941
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-232-201-28

Query Match 80.9%; Score 17.8; DB 4; Length 1941;
Best Local Similarity 90.5%; Pred. No. 30;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 tgcccaggccgaagcctggg 22
||||||| ||||| |||
Db 517 TGCCAGCAGAGGCCAGGG 497

RESULT 9

US-08-785-241-1/c
; Sequence 1, Application US/08785241
; Patent No. 5695963
; GENERAL INFORMATION:
; APPLICANT: McKnight, Steven L.
; APPLICANT: Russell, David W.
; APPLICANT: Tian, Hui
; TITLE OF INVENTION: Endothelial PAS Domain Protein
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 268 BUSH STREET, SUITE 3200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/785,241
; FILING DATE: 17-JAN-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UTS0:1229
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 343-4341
; TELEFAX: (415) 343-4342
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2816 base pairs

; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-785-241-1

Query Match 80.9%; Score 17.8; DB 1; Length 2816;
Best Local Similarity 90.5%; Pred. No. 30;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ctgccaggccgaaggcctgg 21
||||||| |||||
Db 2784 CTGCCAGGTAGAAGGCGCTGG 2764

RESULT 10

US-09-232-200-64/c
; Sequence 64, Application US/09232200A
; Patent No. 6288213
; GENERAL INFORMATION:
; APPLICANT: Stahl, Andreas
; APPLICANT: Hirsch, David J.
; APPLICANT: Lodish, Harvey F.
; APPLICANT: Gimeno, Ruth E.
; APPLICANT: Tartaglia, Louis A.
; TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
; FILE REFERENCE: WH197-21p3MB
; CURRENT APPLICATION NUMBER: US/09/232,200A
; CURRENT FILING DATE: 1999-01-14
; EARLIER APPLICATION NUMBER: 60/071,374
; EARLIER FILING DATE: 1998-01-15
; EARLIER APPLICATION NUMBER: 60/093,491
; EARLIER FILING DATE: 1998-07-20
; EARLIER APPLICATION NUMBER: 60/110,941
; EARLIER FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 64
; LENGTH: 3217
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-232-200-64

Query Match 80.9%; Score 17.8; DB 4; Length 3217;
Best Local Similarity 90.5%; Pred. No. 30;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 tgcccagccgaaggcctgg 22
||||||| |||||
Db 517 TGCCAGGCGAGAGGCCAGGG 497

RESULT 11

US-09-232-197-64/c
; Sequence 64, Application US/09232197A
; Patent No. 6300096
; GENERAL INFORMATION:
; APPLICANT: Stahl, Andreas
; APPLICANT: Hirsch, David J.
; APPLICANT: Lodish, Harvey F.
; APPLICANT: Gimeno, Ruth E.
; APPLICANT: Tartaglia, Louis A.
; TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
; FILE REFERENCE: WH197-21p3MA
; CURRENT APPLICATION NUMBER: US/09/232,197A
; CURRENT FILING DATE: 1999-01-14
; EARLIER APPLICATION NUMBER: 60/071,374
; EARLIER FILING DATE: 1998-01-15
; EARLIER APPLICATION NUMBER: 60/093,491
; EARLIER FILING DATE: 1998-07-20
; EARLIER APPLICATION NUMBER: 60/110,941

; EARLIER FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 64
; LENGTH: 3217
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-232-197-64

Query Match 80.9%; Score 17.8; DB 4; Length 3217;
Best Local Similarity 90.5%; Pred. No. 30;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 tgcccagccgaaggcctgg 22
||||||| |||||
Db 517 TGCCAGGCGAGAGGCCAGGG 497

RESULT 12

US-09-232-201-64/c
; Sequence 64, Application US/09232201A
; Patent No. 6348321
; GENERAL INFORMATION:
; APPLICANT: Stahl, Andreas
; APPLICANT: Hirsch, David J.
; APPLICANT: Lodish, Harvey F.
; APPLICANT: Gimeno, Ruth E.
; APPLICANT: Tartaglia, Louis A.
; TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
; FILE REFERENCE: WH197-21p3MC
; CURRENT APPLICATION NUMBER: US/09/232,201A
; CURRENT FILING DATE: 1999-01-14
; EARLIER APPLICATION NUMBER: 60/071,374
; EARLIER FILING DATE: 1998-01-15
; EARLIER APPLICATION NUMBER: 60/093,491
; EARLIER FILING DATE: 1998-07-20
; EARLIER APPLICATION NUMBER: 60/110,941
; EARLIER FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 64
; LENGTH: 3217
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-232-201-64

Query Match 80.9%; Score 17.8; DB 4; Length 3217;
Best Local Similarity 90.5%; Pred. No. 30;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 tgcccagccgaaggcctgg 22
||||||| |||||
Db 517 TGCCAGGCGAGAGGCCAGGG 497

RESULT 13

US-09-232-200-46/c
; Sequence 46, Application US/09232200A
; Patent No. 6288213
; GENERAL INFORMATION:
; APPLICANT: Stahl, Andreas
; APPLICANT: Hirsch, David J.
; APPLICANT: Lodish, Harvey F.
; APPLICANT: Gimeno, Ruth E.
; APPLICANT: Tartaglia, Louis A.
; TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
; FILE REFERENCE: WH197-21p3MB
; CURRENT APPLICATION NUMBER: US/09/232,200A
; CURRENT FILING DATE: 1999-01-14
; EARLIER APPLICATION NUMBER: 60/071,374
; EARLIER FILING DATE: 1998-01-15

```
; EARLIER APPLICATION NUMBER: 60/093,491
; EARLIER FILING DATE: 1998-07-20
; EARLIER APPLICATION NUMBER: 60/110,941
; EARLIER FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 46
; LENGTH: 3694
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-232-200-46
```

```
Query Match      80.9%; Score 17.8; DB 4; Length 3694;
Best Local Similarity 90.5%; Pred. No. 30;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 2 tgcccaggccgaagcctggg 22
||||| ||||| ||||| |||
Db 691 TGCCAGGCAGAGGCCAGGG 671
```

```
RESULT 14
US-09-232-197-46/c
; Sequence 46, Application US/09232197A
; Patent No. 6300096
; GENERAL INFORMATION:
; APPLICANT: Stahl, Andreas
; APPLICANT: Hirsch, David J.
; APPLICANT: Lodish, Harvey F.
; APPLICANT: Gimeno, Ruth E.
; APPLICANT: Tartaglia, Louis A.
; TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
; FILE REFERENCE: WHI97-21p3MA
; CURRENT APPLICATION NUMBER: US/09/232,197A
; CURRENT FILING DATE: 1999-01-14
; EARLIER APPLICATION NUMBER: 60/071,374
; EARLIER FILING DATE: 1998-01-15
; EARLIER APPLICATION NUMBER: 60/093,491
; EARLIER FILING DATE: 1998-07-20
; EARLIER APPLICATION NUMBER: 60/110,941
; EARLIER FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 46
; LENGTH: 3694
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-232-197-46
```

```
Query Match      80.9%; Score 17.8; DB 4; Length 3694;
Best Local Similarity 90.5%; Pred. No. 30;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 2 tgcccaggccgaagcctggg 22
||||| ||||| ||||| |||
Db 691 TGCCAGGCAGAGGCCAGGG 671
```

```
RESULT 15
US-09-232-201-46/c
; Sequence 46, Application US/09232201A
; Patent No. 6348321
; GENERAL INFORMATION:
; APPLICANT: Stahl, Andreas
; APPLICANT: Hirsch, David J.
; APPLICANT: Lodish, Harvey F.
; APPLICANT: Gimeno, Ruth E.
; APPLICANT: Tartaglia, Louis A.
; TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
; FILE REFERENCE: WHI97-21p3MC
; CURRENT APPLICATION NUMBER: US/09/232,201A
```

```
; CURRENT FILING DATE: 1999-01-14
; EARLIER APPLICATION NUMBER: 60/071,374
; EARLIER FILING DATE: 1998-01-15
; EARLIER APPLICATION NUMBER: 60/093,491
; EARLIER FILING DATE: 1998-07-20
; EARLIER APPLICATION NUMBER: 60/110,941
; EARLIER FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 46
; LENGTH: 3694
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-232-201-46
```

```
Query Match      80.9%; Score 17.8; DB 4; Length 3694;
Best Local Similarity 90.5%; Pred. No. 30;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 2 tgcccaggccgaagcctggg 22
||||| ||||| ||||| |||
Db 691 TGCCAGGCAGAGGCCAGGG 671
```

Search completed: June 19, 2002, 15:47:08
Job time: 13686 sec

Thu Jun 20 06:56:45 2002

us-09-462-955b-1_copy_655_676.rni

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2002, 11:59:02 ; Search time 7489.97 Seconds
(without alignments)
39.644 Million cell updates/sec

Title: US-09-462-955B-1_COPY_655_676

Perfect score: 22

Sequence: 1 ctgcccagggccgaagcctggg 22

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- EST.*
- 1: em_estba.*
 - 2: em_esthum.*
 - 3: em_estin.*
 - 4: em_estmu.*
 - 5: em_estov.*
 - 6: em_estpl.*
 - 7: em_estro.*
 - 8: em_htc.*
 - 9: gb_est1.*
 - 10: gb_est2.*
 - 11: gb_htc.*
 - 12: gb_gss.*
 - 13: em_gss_hum.*
 - 14: em_gss_inv.*
 - 15: em_gss_pln.*
 - 16: em_gss_vrt.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query %	Length	DB	ID	Description
C 1	19.4	88.2	935	10	BG327577	BG327577	60246566F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4564192 5', mRNA sequence.
C 2	18.8	85.5	176	10	BE860673	BE860673	UI-M-AH1-
C 3	18.8	85.5	238	9	BB147701	BB147701	BB147701 BB147701
C 4	18.8	85.5	393	9	BB284196	BB284196	BB284196 BB284196
C 5	18.8	85.5	555	9	BB239924	BB239924	BB239924 BB239924
C 6	18.8	85.5	608	12	BB043080	BB043080	RPCI-24-3
C 7	18.8	85.5	654	10	BB101919	BB101919	602887615
C 8	18.8	85.5	959	10	BB119646	BB119646	601757514
C 9	18.4	83.6	291	10	B1032357	B1032357	CM3-NN024
C 10	18.4	83.6	377	9	BB813476	BB813476	BB813476 BB813476
C 11	18.4	83.6	405	9	AW522130	AW522130	UI-R-BJOp
C 12	18.4	83.6	438	9	AA158436	AA158436	zo59h11.r
C 13	18.4	83.6	455	9	AW535323	AW535323	L0228C11-
C 14	18.4	83.6	512	9	AI406571	AI406571	EST234857
C 15	18.4	83.6	517	10	BF403937	BF403937	UI-R-CA1-
C 16	18.4	83.6	558	10	BT295549	BT295549	UI-R-BK0-
C 17	17.8	80.9	176	10	B1004403	B1004403	MR4-HN005

C 18	17.8	80.9	243	9	AU058457	AU058457	9	AU058457	AU058457
C 19	17.8	80.9	295	9	AW877400	AW877400	9	AW877400	AW877400
C 20	17.8	80.9	313	9	AW336731	AW336731	9	AW336731	AW336731
C 21	17.8	80.9	349	10	BE385251	BE385251	10	BE385251	BE385251
C 22	17.8	80.9	370	10	BE771160	BE771160	10	BE771160	BE771160
C 23	17.8	80.9	371	10	BE372847	BE372847	10	BE372847	BE372847
C 24	17.8	80.9	372	10	BE771169	BE771169	10	BE771169	BE771169
C 25	17.8	80.9	375	10	BE771169	BE771169	10	BE771169	BE771169
C 26	17.8	80.9	377	10	BI342341	BI342341	10	BI342341	BI342341
C 27	17.8	80.9	383	10	BI404521	BI404521	10	BI404521	BI404521
C 28	17.8	80.9	384	10	BI024756	BI024756	10	BI024756	BI024756
C 29	17.8	80.9	390	10	BF087382	BF087382	10	BF087382	BF087382
C 30	17.8	80.9	392	10	AI520113	AI520113	10	AI520113	AI520113
C 31	17.8	80.9	397	9	AW408256	AW408256	9	AW408256	AW408256
C 32	17.8	80.9	401	9	AW838277	AW838277	9	AW838277	AW838277
C 33	17.8	80.9	409	10	BE897489	BE897489	10	BE897489	BE897489
C 34	17.8	80.9	418	10	BE818183	BE818183	10	BE818183	BE818183
C 35	17.8	80.9	430	9	BE014102	BE014102	9	BE014102	BE014102
C 36	17.8	80.9	433	9	BE013607	BE013607	9	BE013607	BE013607
C 37	17.8	80.9	434	9	AW403812	AW403812	9	AW403812	AW403812
C 38	17.8	80.9	435	10	BG609304	BG609304	10	BG609304	BG609304
C 39	17.8	80.9	440	10	BF078233	BF078233	10	BF078233	BF078233
C 40	17.8	80.9	442	9	AA405771	AA405771	9	AA405771	AA405771
C 41	17.8	80.9	445	10	BI401089	BI401089	10	BI401089	BI401089
C 42	17.8	80.9	454	9	AW416681	AW416681	9	AW416681	AW416681
C 43	17.8	80.9	462	10	BF775369	BF775369	10	BF775369	BF775369
C 44	17.8	80.9	471	10	BF444355	BF444355	10	BF444355	BF444355
C 45	17.8	80.9	479	10	BF443589	BF443589	10	BF443589	BF443589

ALIGNMENTS

RESULT 1
BG327577/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BG327577 935 bp mRNA linear EST 27-FEB-2001
60246566F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4564192 5', mRNA sequence.

BG327577
BG327577.1 GI:13134015
EST.
human.

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 935)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: DCTD/DTP

cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LCM1279 row: d column: 17
High quality sequence stop: 789.
Location/Qualifiers
1. .935
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4564192"
/tissue_type="renal cell adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: kidney; Vector: pOTB7; Site:1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in

FEATURES
source

the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)." 222 a 236 c 297 g 180 t

BASE COUNT
ORIGIN

Query Match 88.2%; Score 19.4; DB 10; Length 935;
Best Local Similarity 95.2%; Pred. No. 5.7e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ctgccagccgagcagcctg 21
|||||
DB 910 CTGCCAGCCGAGGAGCCTGG 890

RESULT 2
LOCUS BE860673 176 bp mRNA linear EST 29-SEP-2000
DEFINITION UI-M-AHI-agv-c-04-0-UI-r1 NIH-BMAP_MCE_N Mus musculus cDNA clone
ACCESSION BE860673
VERSION BE860673.1 GI:10377856
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 176)
AUTHORS Bonaldo, M.F., Lennon, G. and Soares, M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 9704477
COMMENT Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD 20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mEST@mail.nih.gov

cDNA Library Preparation: M.B. Soares Lab Clone distribution: Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It should be noted that Bento Soares is generating a small number of additional specialized non-redundant arrays of BMAP cDNAs whose availability will be considered under appropriate and limited collaborative arrangements
Seq primer: M13 Reverse.
Location/Qualifiers
1. 176
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-AHI-agv-c-04-0-UI"
/clone_lib="NIH-BMAP_MCE_N"
/dev_stage="27-32 days"
/lab_host="pH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The NIH-BMAP_MCE library is a normalized library constructed from mouse cerebellum. The tag is a string of 5 nucleotides present between the Not I site and the oligo-dT track. The library was constructed as described by Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996. Tissue provided by Ms. Annie Novakovich, Zivic-Miller Laboratories."

BASE COUNT 29 a 58 c 40 g 49 t
ORIGIN

Query Match 85.5%; Score 18.8; DB 10; Length 176;
Best Local Similarity 90.9%; Pred. No. 7.3e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctgccagccgagcagcctg 22
|||||
DB 71 CTGCTCAGCCGAGGCCCTGG 92

RESULT 3
LOCUS BB147701 238 bp mRNA linear EST 28-JUN-2000

DEFINITION BB147701 RIKEN full-length enriched, adult female vagina Mus musculus cDNA clone 9930101123 3', mRNA sequence.
ACCESSION BB147701
VERSION BB147701.1 GI:8802638
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 238)
AUTHORS Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomimaga, N., Toya, T., Tsunoda, Y., Watahiki, A., Watanabe, S., Yamamura, T., Yamanaka, I., Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
RIKEN Mouse ESTs (Konno, H., et al.)
Unpublished (2000)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

thermostabilization and thermoactivation of thermostable enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh, M., Kitsuai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.
Location/Qualifiers
1. 238
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="9930101123"
/clone_lib="RIKEN full-length enriched, adult female vagina"
/sex="female"
/tissue_type="vagina"
/dev_stage="adult"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken

FEATURES
source

Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@gsr.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh
 M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630 (2000)
 wagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,
 Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, I., Matsura
 S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and
 Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.
 10 (11), 1757-1771 (2000)
 Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara
 Y. and Hayashizaki, Y.
 Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamanaka, I., Aizawa
 K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and
 Hayashizaki, Y.
 Computational Analysis of Full-Length Mouse cDNAs Compared with
 Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
 Please visit our web site (http://genome.gsc.riken.go.jp/) for
 further details.
 cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.

FEATURES source

Location/Qualifiers
 1. .555
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="A630079H08"
 /clone_lib="RIKEN full-length enriched, 3 days neonate
 thymus"
 /tissue_type="thymus"
 /dev_stage="3 days neonate"
 /lab_host="DH10B"
 /note="Site 1: SalI; Site 2: BamHI; cDNA library was
 prepared and sequenced in Mouse Genome Encyclopedia
 Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in
 RIKEN. Division of Experimental Animal Research in Riken
 contributed to prepare mouse tissues. 1st strand cDNA was
 primed with a primer [5'
 GAGAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTT 3'], cDNA was
 prepared by using trehalose thermo-activated reverse
 transcriptase and subsequently enriched for full-length by
 cap-trapper. cDNA went through one round of normalization
 to Rot = 20.0 and subtraction to Rot = 459.0. Second
 strand cDNA was prepared with the primer adapter of
 sequence [5' GAGAGAGAGATTCGAGTTAAATTAATACCCCCCCCCC
 3']. cDNA was cleaved with XhoI and BamHI. Vector: a
 modified pBluescript KS(+) after bulk excision from Lambda
 FLC I."

BASE COUNT 118 a 155 c 140 g 141 t 1 others

LOCUS B1101919/c
 DEFINITION 602887615F1 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAGE:5042882
 5', mRNA sequence.
 ACCESSION B1101919
 VERSION B1101919.1 GI:14552812
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 654)

Query Match 85.5%; Score 18.8; DB 9; Length 555;
 Best Local Similarity 90.9%; Pred. No. 8.9e+02;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctgcccagggcgaagcctggg 22
 ||||| ||||| ||||| ||||| |||||
 Db 414 CTGCTCAGGCGGAGCCCTGGG 435

RESULT 6
 LOCUS BHO43080/c
 DEFINITION BHO43080.1 GI:14824835
 DNA sequence.
 ACCESSION BHO43080
 VERSION BHO43080.1 GI:14824835
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 608)
 Z hao, S., Niemman, W., Malek, J., Shatsman, S., Akinret, B., Levins, M.,
 Tsegaye, G., Geer, K., Krol, M., Shvartsbeyn, A., Gebregorgis, E.,
 Russell, D., de Jong, P. and Fraser, C.M.
 Mouse BAC End Sequences from Library RPCI-24
 Unpublished (1999)
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@tigr.org
 Clones are derived from the mouse BAC library RPCI-24. For BAC
 library availability, please contact Pieter de Jong
 (pdejong@mail.cho.org). Clones may be purchased from BACPAC
 Resources (http://www.tigr.org/tadb/bac_ends/mouse/bac_end_intro.html
 page: http://www.tigr.org/tadb/bac_ends/mouse/bac_end_intro.html
 Plate: 375 row; D column: 2
 Seq primer: SP6
 Class: BAC ends.
 Location/Qualifiers
 1. .608
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-24-375D2"
 /clone_lib="RPCI-24"
 /sex="Male"
 /cell_type="Spleen/Brain"
 /note="Vector: pTARBAC1; Site 1: BamHI; Site 2: BamHI;
 RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
 library was cloned in the pTARBAC1 cloning vector at the
 BamHI sites using MboI partially digested male C57BL/6J
 DNA."

FEATURES source

Location/Qualifiers
 1. .608
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-24-375D2"
 /clone_lib="RPCI-24"
 /sex="Male"
 /cell_type="Spleen/Brain"
 /note="Vector: pTARBAC1; Site 1: BamHI; Site 2: BamHI;
 RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
 library was cloned in the pTARBAC1 cloning vector at the
 BamHI sites using MboI partially digested male C57BL/6J
 DNA."

BASE COUNT 162 a 144 c 160 g 142 t

ORIGIN

Query Match 85.5%; Score 18.8; DB 12; Length 608;
 Best Local Similarity 90.9%; Pred. No. 9.1e+02;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctgcccagggcgaagcctggg 22
 ||||| ||||| ||||| ||||| |||||
 Db 410 CTGCCAGGCGGAGCCCTGGG 389

RESULT 7 LOCUS B1101919/c

DEFINITION 602887615F1 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAGE:5042882
 5', mRNA sequence.
 ACCESSION B1101919
 VERSION B1101919.1 GI:14552812
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 654)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 Tissue Procurement: Jeffrey E. Green, M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Plate: L1AM1117 row: n column: 03
 High quality sequence stop: 654.
 Location/Qualifiers
 1. 654
 /organism="Mus musculus"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="IMAGE:5042882"
 /clone_lib="NCI_CGAP_Kid14"
 /lab_host="DH10B (TI phage-resistant)"
 /note="Organ: kidney; Vector: pCMV-Sport6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dn. Average insert size 1.75 kb. Constructed by Life Technologies. Note: this is a NCI_CGAP Library. |"
 BASE COUNT 172 a 168 c 181 g 133 t
 ORIGIN

FEATURES

Query Match 85.5%; Score 18.8; DB 10; Length 654;
 Best Local Similarity 90.9%; Pred. No. 9.2e+02;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Location/Qualifiers
 1. 654
 /organism="Mus musculus"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="IMAGE:5042882"
 /clone_lib="NCI_CGAP_Kid14"
 /lab_host="DH10B (TI phage-resistant)"
 /note="Organ: kidney; Vector: pCMV-Sport6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dn. Average insert size 1.75 kb. Constructed by Life Technologies. Note: this is a NCI_CGAP Library. |"
 BASE COUNT 172 a 168 c 181 g 133 t
 ORIGIN

Query Match 85.5%; Score 18.8; DB 10; Length 654;
 Best Local Similarity 90.9%; Pred. No. 9.2e+02;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Location/Qualifiers
 1. 654
 /organism="Mus musculus"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="IMAGE:5042882"
 /clone_lib="NCI_CGAP_Kid14"
 /lab_host="DH10B (TI phage-resistant)"
 /note="Organ: kidney; Vector: pCMV-Sport6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dn. Average insert size 1.75 kb. Constructed by Life Technologies. Note: this is a NCI_CGAP Library. |"
 BASE COUNT 172 a 168 c 181 g 133 t
 ORIGIN

Query Match 85.5%; Score 18.8; DB 10; Length 654;
 Best Local Similarity 90.9%; Pred. No. 9.2e+02;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Location/Qualifiers
 1. 654
 /organism="Mus musculus"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="IMAGE:5042882"
 /clone_lib="NCI_CGAP_Kid14"
 /lab_host="DH10B (TI phage-resistant)"
 /note="Organ: kidney; Vector: pCMV-Sport6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dn. Average insert size 1.75 kb. Constructed by Life Technologies. Note: this is a NCI_CGAP Library. |"
 BASE COUNT 172 a 168 c 181 g 133 t
 ORIGIN

Query Match 85.5%; Score 18.8; DB 10; Length 654;
 Best Local Similarity 90.9%; Pred. No. 9.2e+02;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Location/Qualifiers
 1. 654
 /organism="Mus musculus"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="IMAGE:5042882"
 /clone_lib="NCI_CGAP_Kid14"
 /lab_host="DH10B (TI phage-resistant)"
 /note="Organ: kidney; Vector: pCMV-Sport6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dn. Average insert size 1.75 kb. Constructed by Life Technologies. Note: this is a NCI_CGAP Library. |"
 BASE COUNT 172 a 168 c 181 g 133 t
 ORIGIN

Query Match 85.5%; Score 18.8; DB 10; Length 654;
 Best Local Similarity 90.9%; Pred. No. 9.2e+02;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Location/Qualifiers
 1. 654
 /organism="Mus musculus"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="IMAGE:5042882"
 /clone_lib="NCI_CGAP_Kid14"
 /lab_host="DH10B (TI phage-resistant)"
 /note="Organ: kidney; Vector: pCMV-Sport6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dn. Average insert size 1.75 kb. Constructed by Life Technologies. Note: this is a NCI_CGAP Library. |"
 BASE COUNT 172 a 168 c 181 g 133 t
 ORIGIN

Query Match 85.5%; Score 18.8; DB 10; Length 654;
 Best Local Similarity 90.9%; Pred. No. 9.2e+02;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Location/Qualifiers
 1. 654
 /organism="Mus musculus"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="IMAGE:5042882"
 /clone_lib="NCI_CGAP_Kid14"
 /lab_host="DH10B (TI phage-resistant)"
 /note="Organ: kidney; Vector: pCMV-Sport6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dn. Average insert size 1.75 kb. Constructed by Life Technologies. Note: this is a NCI_CGAP Library. |"
 BASE COUNT 172 a 168 c 181 g 133 t
 ORIGIN

Query Match 85.5%; Score 18.8; DB 10; Length 959;
 Best Local Similarity 90.9%; Pred. No. 9.8e+02;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Location/Qualifiers
 1. 291
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="NN0246"
 /dev_stage="Adult"
 /note="Organ: nervous_system; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
 BASE COUNT 67 a 87 c 80 g 57 t
 ORIGIN

Query Match 85.5%; Score 18.8; DB 10; Length 959;
 Best Local Similarity 90.9%; Pred. No. 9.8e+02;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Location/Qualifiers
 1. 291
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="NN0246"
 /dev_stage="Adult"
 /note="Organ: nervous_system; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
 BASE COUNT 67 a 87 c 80 g 57 t
 ORIGIN

Query Match 85.5%; Score 18.8; DB 10; Length 959;
 Best Local Similarity 90.9%; Pred. No. 9.8e+02;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Location/Qualifiers
 1. 291
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="NN0246"
 /dev_stage="Adult"
 /note="Organ: nervous_system; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
 BASE COUNT 67 a 87 c 80 g 57 t
 ORIGIN

Query Match 85.5%; Score 18.8; DB 10; Length 959;
 Best Local Similarity 90.9%; Pred. No. 9.8e+02;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Location/Qualifiers
 1. 291
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="NN0246"
 /dev_stage="Adult"
 /note="Organ: nervous_system; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
 BASE COUNT 67 a 87 c 80 g 57 t
 ORIGIN

Query Match 85.5%; Score 18.8; DB 10; Length 959;
 Best Local Similarity 90.9%; Pred. No. 9.8e+02;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Location/Qualifiers
 1. 291
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="NN0246"
 /dev_stage="Adult"
 /note="Organ: nervous_system; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
 BASE COUNT 67 a 87 c 80 g 57 t
 ORIGIN

Query Match 85.5%; Score 18.8; DB 10; Length 959;
 Best Local Similarity 90.9%; Pred. No. 9.8e+02;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Location/Qualifiers
 1. 291
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="NN0246"
 /dev_stage="Adult"
 /note="Organ: nervous_system; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
 BASE COUNT 67 a 87 c 80 g 57 t
 ORIGIN

Query Match 85.5%; Score 18.8; DB 10; Length 959;
 Best Local Similarity 90.9%; Pred. No. 9.8e+02;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Location/Qualifiers
 1. 291
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="NN0246"
 /dev_stage="Adult"
 /note="Organ: nervous_system; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
 BASE COUNT 67 a 87 c 80 g 57 t
 ORIGIN

Query Match 85.5%; Score 18.8; DB 10; Length 959;
 Best Local Similarity 90.9%; Pred. No. 9.8e+02;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Location/Qualifiers
 1. 291
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="NN0246"
 /dev_stage="Adult"
 /note="Organ: nervous_system; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
 BASE COUNT 67 a 87 c 80 g 57 t
 ORIGIN

BASE COUNT	90 a	102 c	91 g	94 t
ORIGIN				
Query Match	83.6%	Score 18.4;	DB 10;	Length 291;
Best Local Similarity	95.08;	Pred. No. 1.1e+03;		
Matches 19;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;
Qy	2	tgccagcgcgaagcctggg 21		
Db	62	TGCCAGCGCCGAAGGCGTGG 81		
RESULT 10				
BB813476		377 bp	mrna	linear
LOCUS	BB813476	RIKEN full-length enriched, lung RCB-0558 LLC cDNA Mus		EST 19-NOV-2001
DEFINITION	musculus	cDNA clone G730019E17 3', mRNA sequence.		
ACCESSION	BB813476			
VERSION	BB813476.1	GI:16986105		
KEYWORDS	EST.			
SOURCE	house mouse.			
ORGANISM	Mus musculus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
	1 (bases 1 to 377)			
	Akinura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T., Ishii			
	Hayatsu,N., Hiramoto,K., Hirooka,T., Hirozane,T., Imotani,K., Ishii			
	Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T.,			
	Nakamura,N., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T.,			
	Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,			
	Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa			
	A., Takahashi,F., Takaku-Akashira,S., Tanaka,T., Tomaru,A., Toya,T.			
	, Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.			
	RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.			
	2001)			
TITLE	Unpublished (2001)			
JOURNAL	Contact: Yoshihide Hayashizaki			
COMMENT	Laboratory for Genome Exploration Research Group, RIKEN Genomic			
	Sciences Center (GSC), Yokohama Institute			
	The Institute of Physical and Chemical Research (RIKEN)			
	1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan			
	Tel: 81-45-503-9222			
	Fax: 81-45-503-9216			
	Email: genome-res@gsc.riken.go.jp,			
	URL: http://genome.gsc.riken.go.jp/			
	Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh			
	M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.			
	Normalization and subtraction of cap-trapper-selected cDNAs to			
	prepare full-length cDNA libraries for rapid discovery of new			
	genes. Genome Res. 10 (10), 1617-1630 (2000)			
	wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,			
	Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura			
	S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and			
	Hayashizaki,Y.			
	RIKEN integrated sequence analysis (RISA) system-384-format			
	sequencing pipeline with 384 multicapillary sequencer. Genome Res.			
	10 (11), 1757-1771 (2000)			
	Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara			
	Y. and Hayashizaki,Y.			
	Computer-based methods for the mouse full-length cDNA			
	encyclopedia: real-time sequence clustering for construction of a			
	nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)			
	Please visit our web site (http://genome.gsc.riken.go.jp) for			
	further details.			
	e mouse tissues.			
FEATURES	Location/Qualifiers			
source	1. .377			
	/organism="Mus musculus"			
	/db_xref="taxon:10090"			
	/clone="G730019E17"			
	/clone_lib="RIKEN full-length enriched, lung RCB-0558 LLC			
	cDNA"			
	/tissue_type="lung"			
	/cell_line="RCB-0558 LLC"			
Query Match	83.6%	Score 18.4;	DB 10;	Length 291;
Best Local Similarity	95.08;	Pred. No. 1.1e+03;		
Matches 19;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;
Qy	2	tgccagcgcgaagcctggg 21		
Db	62	TGCCAGCGCCGAAGGCGTGG 81		
RESULT 10				
BB813476		377 bp	mrna	linear
LOCUS	BB813476	RIKEN full-length enriched, lung RCB-0558 LLC cDNA Mus		EST 19-NOV-2001
DEFINITION	musculus	cDNA clone G730019E17 3', mRNA sequence.		
ACCESSION	BB813476			
VERSION	BB813476.1	GI:16986105		
KEYWORDS	EST.			
SOURCE	house mouse.			
ORGANISM	Mus musculus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
	1 (bases 1 to 377)			
	Akinura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T., Ishii			
	Hayatsu,N., Hiramoto,K., Hirooka,T., Hirozane,T., Imotani,K., Ishii			
	Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T.,			
	Nakamura,N., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T.,			
	Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,			
	Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa			
	A., Takahashi,F., Takaku-Akashira,S., Tanaka,T., Tomaru,A., Toya,T.			


```

AI406571/c
LOCUS       EST234857 512 bp  mRNA  linear  EST 09-FEB-1999
DEFINITION  Normalized rat kidney, Bento Soares Rattus sp. cDNA clone
            RKIDY92 3' end, mRNA sequence.
ACCESSION   AI406571
VERSION     AI406571.1 GI:4250075
KEYWORDS    EST.
SOURCE      Rattus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus
REFERENCE   1 (bases 1 to 512)
AUTHORS    Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
            Kerlavage,A.R. and Adams,M.D.
TITLE      Rat Genome Project: Generation of a Rat EST (REEST) Catalog & Rat
            Gene Index
JOURNAL    Unpublished (1998)
COMMENT    Contact: Lee, NH
            The Institute for Genomic Research
            9712, Medical Center Drive, Rockville, MD 20850, USA
            Tel: (301)-838-3529
            Fax: (301)-838-0208
            Email: nhlee@tigr.org
            Seq primer: MJ3-21.
FEATURES   Location/Qualifiers
            source          1..512
                        /organism="Rattus sp."
                        /db_xref="taxon:10118"
                        /clone="RKIDY92"
                        /clone_lib="Normalized rat kidney, Bento Soares"
                        /note="Organ: Kidney; Vector: pT73pac; Site_1: EcoRI;
                        Site_2: NotI"
BASE COUNT  116 a 135 c 137 g 124 t
ORIGIN
1
2
3 gccacggccgaagcctctggg 22
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

```

```

oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to identify it as a clone from the
normalized thalamus library cDNA Library Preparation: M.B. Soares
Lab Clone distribution: clones will be available through research
Genetics (www.resgen.com)
Seq primer: MJ3 Forward
POLYA=Yes.
FEATURES   Location/Qualifiers
            source          1..517
                        /organism="Rattus norvegicus"
                        /strain="Sprague-Dawley"
                        /db_xref="taxon:10116"
                        /clone="UI-R-CAL-bjd-j-19-0-UI"
                        /clone_lib="UI-R-CAL"
                        /lab_host="DH10B (Life Technologies)"
                        /note="Vector: pT73D-Pac (Pharmacia) with a modified
                        polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-CAL
                        library is a subtracted library derived from the following
                        tissues: thalamus, cerebellum, hypothalamus, medulla, pons
                        , midbrain, cerebral cortex, corpus striatum, testis, and
                        hippocampus. For a detailed description of the library
                        from which this clone was derived, please visit our web
                        site at ratest.eng.uiowa.edu. The subtraction has been
                        previously described in (Bonaldo, Lennon and Soares,
                        Genome Research 6:791-806, 1996)
                        TAG_LIB=UI-R-CAL
                        TAG_TISSUE=thalamus
                        TAG_SEQ=GATCG"
BASE COUNT  114 a 134 c 133 g 136 t
ORIGIN
1
2
3 gccacggccgaagcctctggg 22
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

```

Search completed: June 19, 2002, 14:04:08
Job time: /506 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2002, 11:59:02 ; Search time 5904.86 Seconds
(without alignments)
77.967 Million cell updates/sec

Title: US-09-462-955B-1_COPY_655_676
Perfect score: 22
Sequence: 1 ctgccagccgaaggcctggg 22

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pi.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
-----	-----	-----	-----	-----	-----

1	22	100.0	23	6	AR172775	Sequence
2	22	100.0	1291	6	AR172770	Sequence
3	22	100.0	1291	14	CFDCG	M29563 Coconut fol
4	19.4	88.2	108930	9	AL513122	Human DNA
5	19.4	88.2	193050	1	AL646062	Ralstonia
6	19.4	88.2	193168	2	AC011969	Homo sapi
7	19.4	88.2	207392	2	AC006558	Homo sapi
8	19	86.4	15982	1	AE004779	Pseudomon
9	19	86.4	168862	9	AC073464	Homo sapi
10	19	86.4	170237	9	AL137070	Human DNA
11	18.8	85.5	324	6	AX315358	Sequence
12	18.8	85.5	2540	10	BC011256	Mus muscu
13	18.8	85.5	25010	9	AL353092	Human DNA
14	18.8	85.5	176253	2	AP001078	Homo sapi
15	18.8	85.5	193487	2	AP001793	Homo sapi
16	18.8	85.5	210734	2	AC015676	Homo sapi
17	18.8	85.5	212055	2	AP000899	Homo sapi
18	18.4	83.6	4091	1	AF323675	Bradyrhiz
19	18.4	83.6	5162	1	AF151965	Pseudomon
20	18.4	83.6	8660	1	AF153207	Pseudomon
21	17.8	80.9	328	4	CFAQP2X1	Y10638 C.familiari
22	17.8	80.9	496	6	AX015318	Sequence
23	17.8	80.9	834	6	AX166558	Sequence
24	17.8	80.9	1140	10	MFEATP2	AF023257 Mus muscu
25	17.8	80.9	1590	6	E01204	DNA sequence
26	17.8	80.9	1938	6	AR168888	Sequence
27	17.8	80.9	1938	6	AX101337	Sequence
28	17.8	80.9	1941	6	AR168887	Sequence
29	17.8	80.9	1941	6	AX015319	Sequence
30	17.8	80.9	1941	6	AX015323	Sequence
31	17.8	80.9	1941	6	AX101336	Sequence
32	17.8	80.9	1941	10	MM015976	U15976 Mus musculu
33	17.8	80.9	2010	9	AK057960	Homo sapi
34	17.8	80.9	2022	9	AK026727	Homo sapi
35	17.8	80.9	2040	9	BC013149	Homo sapi
36	17.8	80.9	2122	9	D83735	Homo sapien
37	17.8	80.9	2149	9	AK024835	Homo sapi
38	17.8	80.9	2222	6	AX015317	Sequence
39	17.8	80.9	2776	6	AX230581	Sequence 1
40	17.8	80.9	2816	6	I84559	Sequence 1
41	17.8	80.9	2818	9	HS081984	Human endot
42	17.8	80.9	3217	6	AR168902	Sequence
43	17.8	80.9	3217	6	AX101372	Sequence
44	17.8	80.9	3694	6	AR168893	Sequence
45	17.8	80.9	3694	6	AX101354	Sequence

ALIGNMENTS

RESULT	1	AR172775	AR172775	Sequence	6 from patent	US 6303345	23 bp	DNA	linear	PAT 17-DEC-2001
LOCUS	AR172775	Sequence	6 from patent	US 6303345	23 bp	DNA	linear	PAT 17-DEC-2001		
DEFINITION	AR172775	Sequence	6 from patent	US 6303345	23 bp	DNA	linear	PAT 17-DEC-2001		
ACCESSION	AR172775	Sequence	6 from patent	US 6303345	23 bp	DNA	linear	PAT 17-DEC-2001		
VERSION	AR172775.1	GI:17912266								
KEYWORDS	Unknown.									
SOURCE	Unknown.									
ORGANISM	Unknown.									
REFERENCE	1 (bases 1 to 23)									
AUTHORS	Rohde, W., Becker, D., Randles, J.W., Hehn, A. and Salamini, F.									
TITLE	Use of a virus DNA as promoter									
JOURNAL	Patent: US 6303345-A 6 16-OCT-2001;									
FEATURES	Location/Qualifiers									
source	1..23									
BASE COUNT	4 a 8 c 9 g 2 t									
ORIGIN	/organism="unknown"									

Query Match 100.0%; Score 22; DB 6; Length 23;
Best Local Similarity 100.0%; Pred. No. 5.4e+02;

Thu Jun 20 06:56:43 2002

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctgcccagccgaagcctggg 22
 Db 1 CTGCCAGGCCGAGGCGCTGGG 22

RESULT 2
 ARL172770 ARL172770 1291 bp DNA linear PAT 17-DEC-2001
 LOCUS Sequence 1 from patent US 6303345.
 DEFINITION
 ACCESSION ARL172770
 VERSION ARL172770.1 GI:17912261
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 UNCLASSIFIED.

REFERENCE 1 (bases 1 to 1291)
 AUTHORS Rohde, W., Becker, D., Randies, J. W., Hehn, A. and Salamini, F.
 TITLE Use of a virus DNA as promoter
 JOURNAL Patent: US 6303345-A 1 16-OCT-2001;
 FEATURES Location/Qualifiers
 1..1291
 /organism="unknown"

BASE COUNT 336 a 323 c 332 g 300 t
 ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 1291;
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctgcccagccgaagcctggg 22
 Db 1045 CTGCCAGGCCGAGGCGCTGGG 1066

RESULT 3
 CFCDCG 1291 bp ss-DNA circular VRL 02-AUG-1993
 LOCUS Cocoonut foliar decay virus, complete genome.
 DEFINITION
 ACCESSION M29963
 VERSION M29963.1 GI:323306
 KEYWORDS circular; complete genome.
 SOURCE Cocoonut foliar decay virus DNA, clones NMORG, pORG, and pORGSAU3A.
 ORGANISM Cocoonut foliar decay virus
 Viruses; ssDNA viruses; Nanovirus.

REFERENCE 1 (bases 1 to 1291)
 AUTHORS Rohde, W., Randies, J. W., Langridge, P. and Hanold, D.
 TITLE Nucleotide sequence of a circular single-stranded DNA associated with cocoonut foliar decay virus
 JOURNAL Virology 176, 648-651 (1990)
 MEDLINE 90266484

COMMENT Draft entry and printed sequence for [1] kindly submitted by W. Rhode, 15-MAR-1989, for release after publication.

FEATURES
 source 1..1291
 /organism="Cocoonut foliar decay virus"
 /db_xref="taxon:12474"

stem_loop 40..70
 /note="stem-loop structure"

CDS 103..975
 /note="ORF 1"
 /codon_start=1
 /protein_id="AAA42894.1"
 /db_xref="GI:323307"

TRANSLATION="MGSSRRWCFTLNVTETEEAANVRRITSLNVAIVGDEVAPS
 TQGRHLQGIHLTKGRQLGKLTVLGNDRPHLEPTRGSDQNRDYSKERVLLHGVV
 TRPVNRPLARFAEPDELEDEPFGTRRCVVGASVETRWAAENPFPFVHWQ
 LEVLSAIGPADDRITLTCWRDGGDKSVFAKYLGLKPDWFTYTCGGRKDVLYIIE
 DKRNILDPVPCNLEYLNALLECVKNRAFSDDKYEPLSLYGLFDHVLVFWANVLDP
 YLKISRDRIKLWNI"
 314..775

CDS

/note="ORF 2"
 /codon_start=1
 /protein_id="AAA42895.1"
 /db_xref="GI:323308"
 /translation="MTGFTWSPVPTNRIETTVRRNGCFSTESRLVLESKGDHWP
 DLRNLMNSAMKQADTEDALYTELRWNGQDPLKIRSHFHTITGSLKCCCLRSQSRT
 IAQSSGYADETEETGSPCLPNISDSSPTGSHVVEPERTYCTSTRTQNEI"
 complement(422..568)
 /note="ORF 6"
 /codon_start=1
 /protein_id="AAA42896.1"
 /db_xref="GI:323309"
 /translation="MEMGTDFORPILSTPPKLRVQIFGIRGLPGGVHQVPPQIVGP
 IVAF"
 639..797
 /note="ORF 3"
 /codon_start=1
 /protein_id="AAA42897.1"
 /db_xref="GI:323310"
 /translation="MRTRRRREVRVQCISRTQARLVLMHWNQKGRIVPVHVGPKTK
 FNPRTQV"
 complement(823..987)
 /note="ORF 5"
 /codon_start=1
 /protein_id="AAA42898.1"
 /db_xref="GI:323311"
 /translation="MTHTLNIPQFYSVPADFQIIRQDITGKYEMHVEPKITKGFFV
 RTECPVINTF"
 1098..1286
 /note="ORF 4"
 /codon_start=1
 /protein_id="AAA42899.1"
 /db_xref="GI:323312"
 /translation="MNRVMGGPTIKDSIWIRNLLCLOCTOPLSTPSIQVSSLLEKK
 AASLYLPSICFCAIGRLS"

CDS

CDS

CDS

CDS

BASE COUNT 336 a 323 c 332 g 300 t
 ORIGIN

Query Match 100.0%; Score 22; DB 14; Length 1291;
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctgcccagccgaagcctggg 22
 Db 1045 CTGCCAGGCCGAGGCGCTGGG 1066

RESULT 4
 AL513122/c 108930 bp DNA linear PRI 26-JUN-2001
 LOCUS Human DNA sequence from clone RP11-477J21 on chromosome 9, complete
 DEFINITION sequence.

ACCESSION AL513122
 VERSION AL513122.7 GI:14572568
 KEYWORDS HTG.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 108930)
 AUTHORS Laird, G.
 TITLE Direct Submission

JOURNAL Submitted (26-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 requests: clonerequest@sanger.ac.uk

COMMENT

On Jun 27, 2001 this sequence version replaced gi:14270608.
 During difference assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; SW:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at <http://www.sanger.ac.uk/Projects/C.elegans/wormpep> This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr9> RP11-477J21 is from the library RPCI-11.2 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm> VECTOR: pBACE3.6

IMPORTANT: This sequence is not the entire insert of clone RP11-477J21 It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true right end of clone RP11-477J21 is at 108930 in this sequence. The true left end of clone RP11-342H3 is at 29985 in this sequence. The true right end of clone RP11-165P4 is at 100 in this sequence.

FEATURES	Location/Qualifiers	source
	1..108930	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/chromosomes="9"	
	/clone="RP11-477J21"	
	/clone_lib="RPCI-11.2"	
repeat_region	1..187	
repeat_region	/note="MPC21B repeat: matches 6..201 of consensus 605..725	
repeat_region	/note="L2 repeat: matches 2624..2746 of consensus 879..1178	
repeat_region	/note="AluJb repeat: matches 1..309 of consensus 1187..1672	
repeat_region	/note="L1MC4 repeat: matches 7492..7975 of consensus 1673..1972	
repeat_region	/note="AluX repeat: matches 1..304 of consensus 1973..2675	
repeat_region	/note="L1MC4 repeat: matches 6783..7492 of consensus 3053..3202	
repeat_region	/note="MIR repeat: matches 82..248 of consensus 4252..4523	
repeat_region	/note="AluSg repeat: matches 1..283 of consensus 4641..4753	
repeat_region	/note="L2 repeat: matches 2589..2710 of consensus 5008..5134	
repeat_region	/note="L1MB1 repeat: matches 6042..6168 of consensus 5296..5413	
repeat_region	/note="HERVL repeat: matches 4908..5022 of consensus 5414..5637	
repeat_region	/note="MLTIG repeat: matches 1..216 of consensus 5638..5940	
repeat_region	/note="AluSg repeat: matches 1..302 of consensus 5941..6209	
repeat_region	/note="MLTIG repeat: matches 216..510 of consensus 6210..6327	
repeat_region	/note="HERVL repeat: matches 4794..4908 of consensus 6328..6781	
repeat_region	/note="MLTIG repeat: matches 1..516 of consensus 6782..7153	
repeat_region	/note="HERVL repeat: matches 4402..4794 of consensus 7157..7314	
repeat_region	/note="FRAM repeat: matches 1..158 of consensus 8178..8363	
repeat_region	/note="MLTIG repeat: matches 1..200 of consensus 8396..8634	
	/note="MLTIC repeat: matches 220..466 of consensus 8664..8966	
	/note="AluSg repeat: matches 1..308 of consensus 9734..10026	
	/note="AluJb repeat: matches 1..304 of consensus 10611..11123	
	/note="MLT2B repeat: matches 1..448 of consensus 12545..12786	
	/note="MLTIC repeat: matches 1..229 of consensus 13425..13855	
	/note="L1 repeat: matches 4205..4658 of consensus 13905..14075	
	/note="MIR repeat: matches 50..216 of consensus 14560..14653	
	/note="MIR repeat: matches 152..250 of consensus 15446..15938	
	/note="MLTID repeat: matches 1..502 of consensus 16000..16056	
	/note="L1MB1 repeat: matches 6105..6162 of consensus 16057..16331	
	/note="AluSg repeat: matches 1..275 of consensus 16332..17014	
	/note="L1MB1 repeat: matches 5442..6105 of consensus 17094..17197	
	/note="L1MC/D repeat: matches 5704..5810 of consensus 17803..18108	
	/note="AluJo repeat: matches 7..308 of consensus 18212..18370	
	/note="FAM repeat: matches 1..159 of consensus 18400..18716	
	/note="L1MC5 repeat: matches 7610..7925 of consensus 20637..20795	
	/note="AluSg/x repeat: matches 140..291 of consensus 20939..21587	
	/note="L2 repeat: matches 1949..2748 of consensus 21631..21947	
	/note="AluJb repeat: matches 1..299 of consensus 22652..22956	
	/note="AluSg repeat: matches 1..305 of consensus 23007..23286	
	/note="AluSg repeat: matches 5..283 of consensus 23327..23364	
	/note="19 copies 2 mer gt 100% conserved"	
	/note="AluJo repeat: matches 1..193 of consensus 23383..23579	
	/note="AluSg repeat: matches 1..298 of consensus 23639..23950	
	/note="AluSg repeat: matches 1..305 of consensus 24131..24437	
	/note="AluSg repeat: matches 1..305 of consensus 24478..24527	
	/note="L2 repeat: matches 2698..2747 of consensus 24643..24880	
	/note="L2 repeat: matches 2248..2476 of consensus 24954..25299	
	/note="MLT1I repeat: matches 68..410 of consensus 25300..25344	
	/note="L2 repeat: matches 2197..2235 of consensus 25345..25707	
	/note="MLT1A1 repeat: matches 1..365 of consensus 25708..25752	
	/note="L2 repeat: matches 2151..2197 of consensus 27394..27701	
	/note="AluJ repeat: matches 1..309 of consensus 27703..27989	
	/note="AluJb repeat: matches 1..295 of consensus 28105..28317	
	/note="L1MC/D repeat: matches 5617..5822 of consensus 28758..29049	
	/note="L1MA4 repeat: matches 5769..6074 of consensus 29050..29372	
	/note="AluSg repeat: matches 1..311 of consensus 29373..29603	
	/note="L1MA4 repeat: matches 6074..6300 of consensus	

```

repeat_region 29698..29949
/note="MIR repeat: matches 12..262 of consensus"
repeat_region 29964..30044
/note="MIR repeat: matches 127..197 of consensus"
repeat_region 30045..30356
/note="AluX repeat: matches 3..312 of consensus"
repeat_region 30357..30390
/note="MIR repeat: matches 92..127 of consensus"
repeat_region 30668..30932
/note="AluX repeat: matches 40..298 of consensus"
repeat_region 31360..31519
/note="L2 repeat: matches 2540..2710 of consensus"
repeat_region 31530..31778
/note="L2 repeat: matches 2483..2746 of consensus"
repeat_region 31834..32101
/note="L2 repeat: matches 2243..2508 of consensus"
repeat_region 33093..33156
/note="MIR repeat: matches 86..151 of consensus"
repeat_region 33878..34052
/note="MIR repeat: matches 20..199 of consensus"
repeat_region 34281..34362
/note="MIR repeat: matches 98..194 of consensus"
repeat_region 34707..34926
/note="MIR repeat: matches 20..262 of consensus"
repeat_region 35373..35551
/note="L2 repeat: matches 2520..2710 of consensus"
repeat_region 35560..35847
/note="MR63B repeat: matches 1..278 of consensus"
repeat_region 35848..36062
/note="THEIC repeat: matches 166..371 of consensus"
repeat_region 36063..36363

Query Match 88.2%; Score 19.4; DB 9; Length 108930;
Best Local Similarity 95.2%; Pred. No. 3.7e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 ctgccagcggccgaagcctgg 21
||||||| |||||
Db 68466 CTGCCAGGCCAAGGCTGG 68446

RESULT 5
AL646062/c
LOCUS
DEFINITION Ralstonia solanacearum GMI1000 chromosome, complete sequence;
segment 6/19
ACCESSION AL646062 AL646052
VERSION AL646062.1 GI:17427974
KEYWORDS
SOURCE
ORGANISM Ralstonia solanacearum.
Ralstonia solanacearum
Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
Ralstonia.
1 (bases 1 to 193050)
Salanoubat,M., Genin,S., Artiguenave,F., Gouzy,J., Mangenot,S.,
Arlat,M., Billault,A., Brottier,P., Camus,J.C., Cattolico,L.,
Chandler,M., Choise,N., Claudel-Renard,C., Cunnac,S., Demange,N.,
Gaspin,C., Lavie,M., Moisan,A., Robert,C., Saurin,W., Schlex,T.,
Siguiet,P., Thebault,P., Whalen,M., Wincker,P., Levy,M.,
Weissenbach,J. and Boucher,C.A.
Genome sequence of the plant pathogen Ralstonia solanacearum
Unpublished
2 (bases 1 to 193050)
Boucher,C.A.
Direct Submission
Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston
Moleculaire des Interactions Plantes-Microorganismes INRA-CNRS,
BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean
Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France, LMGM CNRS
118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA
URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France,
Laboratoire de Biometrie et Intelligence Artificielle INRA, BP27,
F31326 Castanet-Tolosan Cedex. Laboratoire de Genetique Cellulaire
INRA, BP27, F31326 Castanet-Tolosan Cedex
Christian.Boucher@toulouse.inra.fr
http://sequence.toulouse.inra.fr/R.solanacearum.html.
location/Qualifiers
1..193050
/organism="Ralstonia solanacearum"
/db_xref="GMI1000"
/db_xref="taxon:305"
505..771
/gene="RSc0962"
/note="RS04402"
505..771
/gene="RSc0962"
/function="elements of external origin; phage-related
functions and prophages"
/note="product confidence : putative
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="PUTATIVE BACTERIOPHAGE DNA-BINDING TRANSCRIPTION
REGULATOR PROTEIN"
/protein_id="CAD14664.1"
/db_xref="GI:17427975"
/translation="MKYORAHVAACSQFTPAPVSEATAERFLRLPEVMSACGLS
RSSYGVKSGCEFFQPVRLTPSSVAWVDSEIQWIAARIAARQS"
789..1517
/gene="RSc0963"
/note="RS04401"
789..1517
/gene="RSc0963"
/function="miscellaneous; unknown"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Framed"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="HYPOTHETICAL PROTEIN"
/protein_id="CAD14665.1"
/db_xref="GI:17427976"
/translation="MKONTVAALPENQAALTRHQARGKFEARDEADRIADMQK
HRTAADASETAQQAQAALMRSTKTPQDLRLDKAKARAYATAEDYRAIYAEFE
EAHVDAKIGAGVAKLDENAAVTVLTQYADLMSEAKOMLVPLLRATIAQVRAISGA
TFDNAWEYMNESADAALARMYGVIKQAFKFDRAQDAVLQAAIRPADLDRFIPF
SPGATHVNTLRKAAQOQGRGA"
1811..2032
/gene="RSc0964"
/note="RS04403"
1811..2032
/gene="RSc0964"
/function="elements of external origin; phage-related
functions and prophages"
/note="Product confidence : putative
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="PUTATIVE BACTERIOPHAGE TRANSCRIPTIONAL
ACTIVATOR-RELATED TRANSCRIPTION REGULATOR PROTEIN"
/protein_id="CAD14666.1"
/db_xref="GI:17427977"
/translation="MNCPCGSAVATIRSRPVSRTIRELYCQCSNVLGHTFVSLVEV
VRLTSPSSTPDPEVARQLAGRSEPVAAHV"
2400..2576
gene

```

```

/ gene="RSC0965"
/ note="RS04404"
2400. .2576
/ gene="RSC0965"
/ function="miscellaneous; unknown"
/ note="Product confidence : hypothetical
Gene name confidence : hypothetical
Predicted by Codon usage
Predicted by Homology
Predicted by Framed"
/ codon_start=1
/ transl_table=11
/ evidence=not_experimental
/ product="HYPOTHETICAL PROTEIN"
/ protein_id="CAD14667.1"
/ db_xref="GI:17427978"
/ translation="MLDAPARPEQPAFPQILAVITRALRDVAAPPTDRASLDVAGAAL
LAVAAIAQARRRG"
2569. .2892
/ gene="RSC0966"
/ note="RS04400"
2569. .2892
/ gene="RSC0966"
/ function="miscellaneous; hypothetical/global homology"
/ note="Product confidence : hypothetical
Gene name confidence : hypothetical
Predicted by Codon usage
Predicted by Framed"
/ codon_start=1
/ transl_table=11
/ evidence=not_experimental
/ product="CONSERVED HYPOTHETICAL PROTEIN"
/ protein_id="CAD14668.1"
/ db_xref="GI:17427979"
/ translation="MCKRNHPHWMALRQSLTAQLAQQAQPAQTPFAATLARLETVS
VTEML"
2903. .5704
/ gene="RSC0967"
/ note="RS04399"
2903. .5704
/ gene="RSC0967"
/ function="miscellaneous; unknown"
/ note="Product confidence : hypothetical
Gene name confidence : hypothetical
Predicted by Codon usage
Predicted by Framed"
/ codon_start=1
/ transl_table=11
/ evidence=not_experimental
/ product="HYPOTHETICAL PROTEIN"
/ protein_id="CAD14669.1"
/ db_xref="GI:17427980"
/ translation="WHTQSPALFTWFDQKRDIDLHLAERLDLRRSGAKGNVHSPH
HTRASLSIFERGAWDKSTEGGSCIDLVQHCMPDVAIPMEAAKLLGQWYGIIMP
AAPASALAKSTEEYIAERAQNPAPAVLAGVDEAVSRRAIQTGALGNWNS
PKVAGEAGHGPAACFIVRAMDTRVAVDLRYADPAINGNVKTCQGDKLGHGWT
DARRLHAHTYIVESPINALSVECCHPGTAFAALRGTAIVDKIDWTFURGRVVI
ALDHTDPVNERTPMPLGAAWKLSEALTAADIGAMLVDMQDWEEDINDVLAHGA
BELTARMRLLEAWLIPGMPGGERLSCTRVLPDHDGIYWRVREKDFQYVEKFR
HCATDREGOTSRSEELGDLICARVAGLSRLRVQSLATINGTASDQVETFGVSGVFR
AWRGELAALEGSCDYFEPQKCLYNNRFRSTQAARAVIDAYQATFGNAAAT
LWALGTHLKTVLGPHYLQMAKAGKSKLLESLQATSFQVLSGOMLKTDRHRA
SVSYTSHVPQWDFEKLPAKVLSEIDGLQSTYRFTFVGAALTPLYLMCAVLLAGE
EVDVFSLOSKTCRTSLSDVKOGAILPRNLPPFPYWAOLFASQOPERIDLHRAFDV
VLSLSRAETNDATARNVENYAAIMTAWALLAEFAQIDVEQGFIDHLVEMMHATA
ETDGRLPFWWIMELLSLSELDAGRPHPCWDTIOTDSGEMALFLRPSVMDHLSTA
MHLRKFDPALPKTARVFKQQLQSGVVPQHGGRKLEDEKRIILGRCAHMAAISLPR
LEQRLGYALPAAAYQMAS"
complement(6298. .7473)
/ gene="RSC0968"
/ note="RS04398"

```

complement(6298. .7473)
/ gene="RSC0968"
/ function="elements of external origin; transposon-related functions"
/ note="Product confidence : probable
Gene name confidence : hypothetical
Predicted by Codon usage
Predicted by Homology
Predicted by Framed"
/ codon_start=1
/ transl_table=11
/ evidence=not_experimental
/ product="PROBABLE CP4-LIKE INTEGRASE PROTEIN"
/ protein_id="CAD14670.1"
/ db_xref="GI:17427981"
/ translation="MPLTDLISARNAKPOKPYRLADGLGMYLEVMNPGSKYWRLLKYRI
AGKRRALGVPSPVLAQARTDARRLIAAGTDPQOARKEAKLATTTARTTTFEV
VAREWHEKOSATVPTYATVLSHLETFDPLGARPTETAPOMLATLKKVSRDA
LETVARLOQCSAVFRVAVTGRAHNCTDLKGAFKTAQVQHMARVOANVEPFLIKN
IRGYDGLITRIALQFMALTVFRTKELGAQWSEIDFETEWRIPAERMKRPHEFVP
LSWQAIALLRAAKMTGRQVYFOSPGRTHISNNTILYALRYLGYHSRMTGHGFRGL
ASTMLHEQGFNPVIEROLAHAEKRVRAYNHQAQYLPERRRMQHWADHLDQLAQD"
complement(7649. .7723)
/ gene="tRNA-argW"

Query Match 88.2%; Score 19.4; DB 1; Length 193050;
Best Local Similarity 95.2%; Pred. No. 3.le+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 tgccacgagccgaagcctggg 22
|||||
Db 171301 TGCCAGCGCGAGGCGCTCGG 171281

RESULT 6
AC011969/c
LOCUS
DEFINITION
AC011969
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
1 (bases 1 to 193168)
Homo sapiens chromosome 4, clone RP11-520J8
Unpublished
2 (bases 1 to 193168)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
Baldwin, J., Barna, N., Becker, R., Boguslavsky, L., Boukhgaiter, B.,
Brown, A., Castle, A., Colangelo, M., Collins, S., Collins, S.,
Cooke, P., DeArrellano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,
Ferrelira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D.,
Gallagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J.C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
Lehoczky, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,
McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrim, J.,
Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
Wyman, D., Ye, W.J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (17-OCT-1999), Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 21, 2000 this sequence version replaced gi:7008795.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L2649
 Center clone name: 520-J.8

----- Summary Statistics
 Sequencing vector: M13; M77815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 126832 bases at least Q40
 Consensus quality: 155745 bases at least Q30
 Consensus quality: 175228 bases at least Q20
 Insert size: 186000; agarose-fp
 Insert size: 191968; sum-of-contigs
 Quality coverage: 3.2 in Q20 bases; agarose-fp
 Quality coverage: 3.1 in Q20 bases; sum-of-contigs

*** NOTE: This is a 'working draft' sequence. It currently
 * consists of 13 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1093: contig of 1093 bp in length
 1094 1193: gap of 100 bp
 1194 2376: contig of 1183 bp in length
 2377 2476: gap of 100 bp
 2477 5933: contig of 3457 bp in length
 5934 6033: gap of 100 bp
 6034 9960: contig of 3927 bp in length
 9961 10060: gap of 100 bp
 10061 16555: contig of 6495 bp in length
 16556 16655: gap of 100 bp
 16656 22671: contig of 6016 bp in length
 22672 22771: gap of 100 bp
 22772 31113: contig of 8342 bp in length
 31114 31213: gap of 100 bp
 31214 41484: contig of 10271 bp in length
 41485 41584: gap of 100 bp
 41585 50824: contig of 9240 bp in length
 50825 50924: gap of 100 bp
 50925 69485: contig of 18561 bp in length
 69486 69585: gap of 100 bp
 69586 99210: contig of 29625 bp in length
 99211 99310: gap of 100 bp
 99311 130378: contig of 31068 bp in length
 130379 130478: gap of 100 bp
 130479 193168: contig of 62690 bp in length.

FEATURES

source
 1. 193168
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="4"
 /map="A"
 /clone="RP11-520J8"
 /clone_lib="RPC1-11 Human Male BAC"
 1. 1093
 /note="assembly_fragment"
 1194. 2376
 /note="assembly_fragment"
 2477. 5933
 /note="assembly_fragment"
 6034. 9960
 /note="assembly_fragment"
 10061. 16555
 /note="assembly_fragment"
 16656. 22671
 /note="assembly_fragment"
 22772. 31113

misc_feature
 31214. 41484
 /note="assembly_fragment"
 41585. 50824
 /note="assembly_fragment"
 50925. 69485
 /note="assembly_fragment"
 69586. 99210
 /note="assembly_fragment"
 99311. 130378
 /note="assembly_fragment"
 clone_end:SP6
 vector_side:left
 130479. 193168
 /note="assembly_fragment"
 BASE COUNT 55031 a 40516 c 42876 g 53468 t 1277 others
 ORIGIN

Query Match 88.2%; Score 19.4; DB 2; Length 193168;
 Best Local Similarity 95.2%; Pred. No. 3.1e+02;
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ctgccacgcccgaagcctgg 21
 |||||
 Db 115095 CTGCCACGCGCAAGCCTGG 115075

RESULT 7

AC006558/207392 bp DNA linear HTG 10-FEB-1999
 Homo sapiens chromosome 9 clone hRPK.477_J_21 map 9, *** SEQUENCING
 IN PROGRESS ***; 7 unordered pieces.
 AC006558
 HTG: HTGS_PHASE1.
 human.

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 1 (bases 1 to 207392)

Unpublished
 2 (bases 1 to 207392)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
 Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
 Castle,A., Cerny,J., Collangelo,M., Collins,S., Collymore,A.,
 Cooke,P., DeArelano,K., Depayre,E., Devon,K., Dewar,K.,
 Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
 Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
 Hagoas,B., Hearford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
 Karatas,A., Lehoczy,J., Lieu,C., Locke,K., Macdonald,P.,
 Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
 Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
 Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
 Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
 Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
 Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
 Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.

Direct Submission
 Submitted (10-FEB-1999) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 All repeats were identified using RepeatMasker: Smit, A.F.A. &
 Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 7 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will

* be preserved.
 1 3313: contig of 3313 bp in length
 * gap of unknown length
 * 3314 9632: contig of 6319 bp in length
 * gap of unknown length
 * 9633 24168: contig of 14536 bp in length
 * gap of unknown length
 * 24169 38366: contig of 14198 bp in length
 * gap of unknown length
 * 38367 56172: contig of 17806 bp in length
 * gap of unknown length
 * 56173 87103: contig of 30931 bp in length
 * gap of unknown length
 * 87104 207392: contig of 120289 bp in length.

FEATURES

source

Location/Qualifiers
 1..207392
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="9"
 /map="g"
 /clone="hRPK 477 J-21"
 /clone_lib="RCI-11 human BAC library"
 BASE COUNT 57991 a 44607 c 45345 g 58166 t 1283 others
 ORIGIN

Query Match 88.2%; Score 19.4; DB 2; Length 207392;
 Best Local Similarity 95.2%; Pred. No. 3e+02; 1; Indels 0; Gaps 0;
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ctgcccagccgaaggcctgg 21
 |||||
 Db 70048 CTGCCAGCCAAAGGCGCTGG 70028

RESULT

8

AE004779 15982 bp DNA linear BCT 30-AUG-2000
 LOCUS Pseudomonas aeruginosa PA01, section 340 of 529 of the complete
 DEFINITION genome.
 ACCESSION AE004779 AE004091
 VERSION AE004779.1 GI:9949735
 KEYWORDS
 SOURCE Pseudomonas aeruginosa.
 ORGANISM Pseudomonas aeruginosa
 Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 Pseudomonas.

REFERENCE 1 (bases 1 to 15982)
 AUTHORS Stover,C.K., Pham,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warriner,P.,
 Hickey,M.J., Brinkman,F.S., Hufnagle,W.O., Kowalik,D.J., Lagrou,M.,
 Garber,R.L., Golltry,L., Tolentino,E., Westbrook-Wadman,S., Yuan,Y.,
 Brody,L.L., Coulter,S.N., Folger,K.R., Kas,A., Larbig,K., Lim,R.,
 Smith,K., Spencer,D., Wong,G.K., Wu,Z., and Paulsen,I.T.
 Complete genome sequence of Pseudomonas aeruginosa PA01, an
 opportunistic pathogen
 JOURNAL Nature 406 (6799), 959-964 (2000)
 MEDLINE 2043737
 REFERENCE 2 (bases 1 to 15982)
 AUTHORS Stover,C.K., Pham,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warriner,P.,
 Hickey,M.J., Brinkman,F.S., Hufnagle,W.O., Kowalik,D.J.,
 Lagrou,M., Garber,R.L., Golltry,L., Tolentino,E.,
 Westbrook-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N.,
 Folger,K.R., Kas,A., Larbig,K., Lim,R.M., Smith,K.A., Spencer,D.H.,
 Wong,G.K.-S., Wu,Z., Paulsen,I.T., Reizer,J., Sailer,M.H.,
 Hancock,R.E.W., Lory,S. and Olson,M.V.
 Direct Submission

TITLE Submitted (16-MAY-2000) Department of Medicine and Genetics,
 JOURNAL University of Washington Genome Center, University of Washington,
 Box 352145, Seattle, WA 98195, USA
 Location/Qualifiers
 1..15982
 /organism="Pseudomonas aeruginosa"
 /strain="PA01"

FEATURES

source

gene

gene

CDS

/db_xref="taxon:287"
 163..1002
 /gene="glpF"
 /note="PA3581"
 163..1002
 /gene="glpF"
 /codon_start=1
 /transl_table=11
 /product="glycerol uptake facilitator protein"
 /protein_id="AAG06969.1"
 /db_xref="GI:9949736"
 /translation="MTTAAPTPSLFGQCIAEFLGTALLIFFGTGCAALKVAGASFG
 WEISLWGVGSMALYISAGVSGAHLNPAVSIALWLFAGFGRKLFPFVITAAVGAFC
 AALVYITVSSLTFEFAQNIVRGSDSLASVSTYFPHALSVGQAFIVEVITA
 IIMAVIMALTDDGNLPRGLAPLIGLLIIVIGSAMGLTGFAMNPAQDFGPKLMTY
 LAGWGIPTAGTGREIPYFLVPIFAPILGACILGAGGYRVLIARHLPSAAAPAEPEKV
 RAS"
 1042..2559
 /gene="glpK"
 /note="PA3582"
 1042..2559
 /gene="glpK"
 /codon_start=1
 /transl_table=11
 /product="glycerol kinase"
 /protein_id="AAG06970.1"
 /db_xref="GI:9949737"
 /translation="MTDKHKKYVVALDQGTSSRAIVFDRDANVVSQAQREPAQFYP
 QAGVHDHPMEIATQSTLVEALQAQSIHDOVAAGITNORETTVMWDRHSGRPIH
 NAIWQCRRAAICAOLKROGLEDYIRETTGLVTDYFSGTKLWILDNVEGARERAR
 NGDLLEGTIDFWLLWKLTEGKVHVDTYNASRTWLFNHSRDMARLEVLIDIPRSL
 PEVNSSEVIGNARIIGVGGEPLPIAGTQQAALFGQMCVEPGAKNTYGTGCFLL
 MHTGDKAVKSTHGLTTIACGPRGEVYALEGAVFNGGSGVQWLRLKTLSTAYOT
 YFATKVKDSNGVLYPAFTGLGAPYWPYARGAVFLTRGVKADHLIRATLSTAYOT
 RDVLDMQDAGERLRALRVDDGAVANNFLMFOADILGTTRVPRVWRETATLGAAYL
 AGLACFWSSLDLKSXAVIERVPECEDEPRREKLYAGWKKAVERTRGWDDEL"
 2765..3520
 /gene="glpR"
 /note="PA3583"
 2765..3520
 /gene="glpR"
 /codon_start=1
 /transl_table=11
 /product="glycerol-3-phosphate regulon repressor"
 /protein_id="AAG06971.1"
 /db_xref="GI:9949738"
 /translation="MNLPPRQQTIIIEVRERGYLSIEEQAQFEAVTPOTLRDDINOLA
 EQGLIRRYHGGAAVDSSIENTAVTMRADQMRDEKQRTAEAVASILVDPNASLFINIGTT
 TEALARALLNHRNLKIITNLHVAATLSAKEDFVLVAGCTVRSDDGIVGQAAVDIFIQ
 QFVDFALVIGSIDEDGSLDDEFYQEVRYSOAILDNAROVFLAADSCKFGRNAVRL
 GPISLVNRVSDSPPPAAIARLMAQHKVHLDLV"
 3798..5336
 /gene="glpD"
 /note="PA3584"
 3798..5336
 /gene="glpD"
 /codon_start=1
 /transl_table=11
 /product="glycerol-3-phosphate dehydrogenase"
 /protein_id="AAG06972.1"
 /db_xref="GI:9949739"
 /translation="MSQAHTPSAPLAEVVDVAVVGGINGVGTAAADAAGRGLSVFLICE
 QHDLAQTSSAKSLIHGRLYRLEHFEFLVRELAERAEVLLAKPHIVKPLKFLVPLP
 RPHLRPAWIRAGFLFDHLGKREKLPASGLRFTGSSPLKAEIRRGFFYSQCAVDPA
 RLVLVNAISAREHGAHVHTRCVRSSRRSGKLWHLHLERSLQSDGSLYIRAPVLDQFT
 WVARFTQDGLKQSPYIGIRLIQSGHIIIVPKLYEGEHAYILQNDREIRIVRAIPVLDQFT
 MICTDREYQDPAKVAIBEEFTAYLLQVNAHFQKQAAADTLHSAFAGVRLPQDDES
 DEPSAITRYDTLSAGNEPPLLSVFGKLTTRYKLAESALTQLOFPFANLPQAWTA
 KAPLPGEQMSVEALTEQIANRYAMLDRELALRWARTYGTWRWLLDGVNREADIGE
 HGGGLYAREVDYLCKHEWAQDAEDILWRRSKLGFLSPSQOQVRLGQYLQSEBHPHRP
 VHAA"
 5469..5798
 /gene="glpM"

Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@watson.wustl.edu
----- Summary Statistics
Center project name: H_NH0134N21

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/qsc>

SOURCE INFORMATION:

The rPCI-11 human BAC library was made from the blood of one male donor, as described by Oseogawa, K., Woon, P. Y., Zhao, B., Prengden, E., Tateno, M., Catane, J. J. and de Jong, P. J. (1998) an in-house approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the right is RP11-71B7. Actual start of this clone is at base position 1 of RP11-134N21; actual end is at base position 168862 of RP11-134N21.

Data from AC068165 and AC092432 was used to finish this clone, AC073464.

FEATURES
source

[illegible]

repeat_region	/rpt_family="MER1_type" 11088. .11205 /rpt_family="MER1_type" 11205. .11233 /rpt_family="(TTTTTA)n" 11206. .11523 /rpt_family="Alu" 11524. .11584 /rpt_family="MER1_type" 11912. .12163 /notes="similar to EST BE465434 (NID:g95111209) hw20c01.x1"
repeat_region	12120. .12163 /rpt_family="MER1_type" 12164. .12457 /rpt_family="Alu" 12472. .12521 /rpt_family="(CAAAA)n" 12517. .12644 /rpt_family="MER1_type" 12585. .13001 /notes="similar to EST N22811 (NID:g1136961) yw41f03.s1"
repeat_region	13467. .13785 /rpt_family="MER2_type" 13805. .13825 /rpt_family="AT_rich" 14487. .14510 /rpt_family="(TTTTA)n" 14864. .14891 /rpt_family="AT_rich" 15245. .15496 /rpt_family="L2" 15510. .15601 /rpt_family="MIR" 15712. .15734 /rpt_family="AT_rich" 16076. .16157 /rpt_family="CT_rich" 17187. .17215 /rpt_family="AT_rich" 17724. .17755 /rpt_family="AT_rich" 18261. .18519 /rpt_family="L2" 18531. .18551 /rpt_family="AT_rich" 19293. .19370 /rpt_family="MER1_type" 19343. .19370 /rpt_family="AT_rich" 19618. .19644 /rpt_family="(T)n" 22636. .22658 /rpt_family="AT_rich" 23293. .23376 /rpt_family="CR1" 23495. .23590 /rpt_family="MER1_type" 25273. .25602 /note="similar to EST AV724297 (NID:g10828544) "
repeat_region	26108. .26129 /rpt_family="AT_rich" 26385. .26616 /rpt_family="MER1_type" 26821. .26929 /rpt_family="Achobo" 26951. .27237 /rpt_family="Achobo" 27520. .27880 /rpt_family="ERVL" 27912. .28072 /rpt_family="ERVL" 28986. .29431 /rpt_family="L1"

Thu Jun 20 06:56:43 2002

repeat_region	29506..29537
repeat_region	/rpt_family="AT_rich"
repeat_region	29884..30783
repeat_region	/rpt_family="L1"
repeat_region	30871..31273
repeat_region	/rpt_family="L1"
repeat_region	31261..31285
repeat_region	/rpt_family="AT_rich"
repeat_region	31807..31839
repeat_region	/rpt_family="AT_rich"
repeat_region	32015..32038

Query Match 86.4%: Score 19: DB 9: Length 168862:

Query Match
Best Local Similarity
100.0%: Pred. No 4.6e+02;
30.4%: score 13, 255,

BEST LOCAL SIMILARITY	100.00%	PIDC: NO.	4.00(0.02)
MATCHES	19:	Conservative	0; Mismatches 0;
		Indels	0; Gaps 0;

0v 3 qcccaqccqaaqqcctq 21

Q1 5 gccacggccgacggcccgcccgccg 21
|||||
Db 65377 GCCCAGGCCGGAAGGCTGG 65395

RESULT 10

RESULT TO
AL137070/C

Locus AL137070/C

DEFINITION

2 calponin 2 (CNN2) pseudogenes, a gene for a novel protein similar to aquaporin 7 (AQP7), part of gene for a novel protein similar to methyltetrahydrofolate dehydrogenase (NADP+ dependent), methyltetrahydrofolate cyclohydrolase, formyltetrahydrofolate synthetase (MTHFR1), and a Cdc12L complete sequence.

ACCESSION
AL137070
SYNCHETAS

ACCESSION	AL137070	GI:9367917
VERSION	AL137070.9	

VERSION
KEYWORDS
AAT37070.3 31.507517
HTG: calponin; CNN2; CpG island.

SOURCE human.

ORGANISM	Homo sapiens
ORGANISM	Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
1 (bases 1 to 170237)

AUTHORS Skuce, C.

TITLE	Direct Submission
1. Abstract	1. Abstract
2. Introduction	2. Introduction
3. Methods	3. Methods
4. Results	4. Results
5. Discussion	5. Discussion
6. Conclusion	6. Conclusion
7. References	7. References
8. Appendix	8. Appendix
9. Supplementary Materials	9. Supplementary Materials
10. Other	10. Other

JOURNAL Submitted (04-DEC-2001) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: hamquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT On Jul 22, 2000 this sequence version replaced gi:8977609.

On the 227,2000 assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

Only a small overlap as described above. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:

numbers given in the text and the following abbreviations: Emn: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at <http://www.sanger.ac.uk/projects/C-elegans/wormpep> This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr9>

<http://www.sanger.ac.uk/ysr/ymr/> This sequence is the entire insert of clone RP11-251017. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. RP11-251017 is from the library RPCI-11.1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm> VECTOR: pBACe3.6.

FEATURES

source

Location/Qualifiers

1. .170237

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-251017"
/clone_lib="RPC1-11.1"
3..433
/note="match: GSS: Em:AQ313358"
5..364
/note="match: GSS: Em:AQ884352"
7..785
/note="match: GSS: Em:AQ741764"
7..439
/note="match: GSS: Em:AQ182496"
92..384
/note="AluJo repeat: matches 1..288 of consensus"
complement(527..1042)
/note="match: GSS: Em:AQ174050"
1467..1634
/note="2 copies 84 mer 81% conserved"
2090..2613
/note="LIPB2 repeat: matches 5614..6155 of consensus"
complement(2326..3058)
/note="match: GSS: Em:AQ078000"
/note="match: GSS: Em:AQ749938"
/note="match: GSS: Em:AQ530925"
complement(2586..3082)
/note="match: GSS: Em:AQ435304"
complement(2606..3087)
/note="match: GSS: Em:AQ416033"
complement(2638..3058)
/note="match: GSS: Em:AQ057987"
complement(2642..3082)
/note="match: GSS: Em:AQ200338"
complement(2670..3084)
/note="match: GSS: Em:AQ494022"
complement(2676..3081)
/note="match: GSS: Em:AQ719378"
complement(2677..3085)
/note="match: GSS: Em:AQ410685"
3059..3150
/note="match: GSS: Em:AQ038020"
3068..3150
/note="match: GSS: Em:AQ239460"
3068..3362
/note="match: GSS: Em:AQ111808"
3088..3353
/note="match: GSS: Em:AQ810980"
3089..3264
/note="match: GSS: Em:AQ831431"
3101..3563
/note="match: GSS: Em:AQ559024"
3101..3479
/note="match: GSS: Em:AQ493601"
3106..3531
/note="match: GSS: Em:AQ033506"
3115..3699
/note="match: GSS: Em:AQ045861"
3155..3616
/note="match: GSS: Em:AQ239460"
3155..3700
/note="match: GSS: Em:AQ525695"
3155..3388
/note="match: GSS: Em:AQ038020"
3266..3538
/note="match: GSS: Em:AQ831431"
3348..3735
/note="MER70A repeat: matches 166..564 of consensus"
3355..3572
/note="match: GSS: Em:AQ810980"
4107..4340
/note="L1M4 repeat: matches 5352..5586 of consensus"

```



```

repeat_region 4398..4519
/note="MSTB repeat: matches 1..121 of consensus"
repeat_region 4520..4733
/note="MER30 repeat: matches 1..230 of consensus"
misc_feature complement(4614..5113)
/note="match: GSS: Em:AQ547317"
misc_feature complement(4616..5130)
/note="match: GSS: Em:AQ380868"
misc_feature complement(4616..5127)
/note="match: GSS: Em:AQ748099"
misc_feature complement(4616..5124)
/note="match: GSS: Em:AQ682825"
misc_feature complement(4661..5373)
/note="match: GSS: Em:AQ194341"
misc_feature 4734..5306
/note="match: GSS: Em:AQ269571"
repeat_region 4734..5035
/note="MSTB repeat: matches 121..426 of consensus"
misc_feature complement(4775..5372)
/note="match: GSS: Em:AQ418853"
misc_feature complement(4835..5359)
/note="match: GSS: Em:AQ664754"
misc_feature complement(4837..5369)
/note="match: GSS: Em:AQ454594"
misc_feature complement(4860..5375)
/note="match: GSS: Em:AQ346755"
misc_feature complement(4865..5375)
/note="match: GSS: Em:AQ549101"
misc_feature complement(4885..5367)
/note="match: GSS: Em:AQ693755"
misc_feature 5131..5586
/note="match: GSS: Em:AQ409721"
misc_feature 5131..5574
/note="match: GSS: Em:AQ798688"
misc_feature 5131..5633
/note="match: GSS: Em:AQ693759"
misc_feature 5170..5513
/note="match: GSS: Em:AQ282181"
misc_feature 5381..5493
/note="match: GSS: Em:AQ808738"
misc_feature 5386..5658
/note="match: GSS: Em:AQ426687"
misc_feature 5389..5789
/note="match: GSS: Em:AQ520858"
misc_feature 5393..5859
/note="match: GSS: Em:AQ180925"
misc_feature 5404..5659
/note="match: GSS: Em:AQ764635"
misc_feature 5431..5593
/note="match: GSS: Em:B75305"
misc_feature 5450..5792
/note="match: GSS: Em:B82392"
misc_feature 5497..5900
/note="match: GSS: Em:AQ808738"
repeat_region 5775..6068
/note="AluJo repeat: matches 1..282 of consensus"
misc_feature 5793..5946
/note="match: GSS: Em:AQ520858"
misc_feature complement(6293..6887)
/note="match: GSS: Em:B65313"
misc_feature complement(6419..6895)
/note="match: GSS: Em:AQ232792"
misc_feature complement(6420..6876)
/note="match: GSS: Em:AQ178179"
misc_feature complement(join(6671..7023,7998)..8098))
/note="match: GSS: Em:AQ278128"
misc_feature complement(join(6703..7023,7998)..8166))
/note="match: GSS: Em:AQ571063"
misc_feature complement(6772..7032)
/note="match: GSS: Em:AQ512891"
misc_feature 6875..7321
/note="match: GSS: Em:AQ759250"
misc_feature 6902..7238

```

```

misc_feature /note="match: GSS: Em:AQ770277"
6987..7346
repeat_region /note="match: GSS: Em:AQ228275"
7030..7997

Query Match 86.4%; Score 19; DB 9; Length 170237;
Best Local Similarity 100.0%; Pred. NO. 4.6e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 gccagagccgaagcctgg 21
|||||
Db 26857 GCCAGGCCGAGGCCTGG 26839
|||||

RESULT 11
AX315358/c
LOCUS AX315358 324 bp DNA linear PAT 14-DEC-2001
DEFINITION Sequence 8343 from Patent WO0190366.
ACCESSION AX315358
VERSION AX315358.1 GI:17898914
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Leach,M.D. and Shinkets,R.A.
TITLE Human polynucleotides and polypeptides encoded thereby
JOURNAL Patent: WO 0190366-A 8343 29-NOV-2001;
Curagen Corporation (US)
FEATURES
Source
1..324
Location/Qualifiers
BASE COUNT 82 a 84 c 89 g 69 t
ORIGIN
/note="match: GSS: Em:AQ770277"
6987..7346
repeat_region /note="match: GSS: Em:AQ228275"
7030..7997

Query Match 85.5%; Score 18.8; DB 6; Length 324;
Best Local Similarity 90.9%; Pred. NO. 3.8e+03;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctgcccagccgaagcctggg 22
|||||
Db 71 CTGCCAGGCCGAGTGCCTGGG 50
|||||

RESULT 12
BC011256/c
LOCUS BC011256 2540 bp mRNA linear ROD 30-JUL-2001
DEFINITION Mus musculus, Similar to plasminogen activator, tissue, clone
MGC:18508 IMAGE:4038404, mRNA, complete cds.
ACCESSION BC011256
VERSION BC011256.1 GI:15030027
KEYWORDS MGC.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 2540)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (25-JUL-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome

```

Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: villalon@bcm.tmc.edu.
Villalon, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia, A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W., Muzny, D.M., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK Plate: 24 Row: d Column: 24
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 202109.

FEATURES
Location/Qualifiers
1..2540
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="MGC:18508 IMAGE:4038404"
/tissue_type="Mammary tumor. WAP-TGF alpha model. 7 months old, gross tissue."
/clone_lib="NCI_CGAP_Mam5"
/lab_host="DH108"
/note="Vector: pCMV-SPORT6"
108..1787
/codon_start=1
/product="Similar to plasminogen activator, tissue"
/protein_id="AAH11256.1"
/db_xref="GI:15030028"
/translation="MKRELICVLLGLAPLPDQGIHGRFRGARSYRATCRDEPTQ
TTYQHOSLRLPMLRSVEYRCNSGLVQCHSVPRSCSEPCFNGTCQOALYFSD
FVQCQDPGFKGCDIDTRATCEOGIYRTWTSAESGAEICINWNSVLKYPNA
RRNAIKLGNHNYCRNDRDLKPMCYVFKAGKYTFECSTPACPKSGEDCYGKG
VYRTHSLTSSQALPWSIVLMGKSTAWNTNSQALGHRHNYCRNPDGADAPWC
HVMKRLKWEYCDMSPCSTGRLQYRPFRIKGLYDITSHPMQAEIFVANKRSP
GERFLCGVLISCVWLSAHCFLEFPNHLKVLGRTYRVVPGEEQTFEIEKTIY
HEFFDDYNDIALQLRSQKQCAESSVGTACLPDNLQLPDTECELSGYGKH
EASPFSDRLKHAHRLXPSSRCTSQHLFNTVNNMLCAGDTRSGNQDLHDACOG
DSGGPLVCMINKOMTLGTIISWGLGQKQDVPVYTKVTNYLDWIHDMKQ"

BASE COUNT
ORIGIN
699 a 619 c 644 g 578 t

Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:
Em: EMBL; SW: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr20>
IMPORTANT: This sequence is not the entire insert of clone RPI-180I13 It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.
The true left end of clone RFS-836N17 is at 24911 in this sequence. The true right end of clone RPI-3100I3 is at 100 in this sequence. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. RPI-180I13 is from the library RPI-1 constructed by the group of Pieter de Jong. For further details see
<http://www.chori.org/bacpac/home.htm>
VECTOR: pCYPAC2.
Location/Qualifiers
1..25010
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/clone="RPI-180I13"
/clone_lib="RPI-1"
1..43
/note="AluJo/FLAM repeat: matches 76..118 of consensus"
complement(39..567)
/note="match: GSS: Em:AQ386884"
58..175
/note="5S repeat: matches 1..119 of consensus"
285..435
/note="MER20 repeat: matches 26..218 of consensus"
510..586
/note="L2 repeat: matches 1966..2045 of consensus"
692..906
/note="match: GSS: Em:AQ486790"
772..843
/note="2 copies 36 mer 93% conserved"
1576..2453
/note="CpG island"
/evidence="not_experimental"
join(1846..1956,21132..21252,22789..22831,23186..23288,24093..>24191)
/gene="HCK"
/note="match: CDNAS: Em:M16591 Em:X62345 Em:S74141 Em:J03023 Em:M16592 Em:M83666 Em:Y00487 Em:X60380 Em:J03579 Em:X67786 Em:AF000300 Em:AF000301 Em:AF000302 Em:M17031 Em:X52822 Em:X57018 Em:X15345 Em:M27454 Em:M57290 Em:X57684 Em:M85043 Em:M19722
match: ESTs: Em:AW307786"
/product="dl180i13.1 (hemopoietic cell kinase)"
/evidence="not_experimental"
1846..24191
/gene="HCK"
complement(2102..2519)
/note="match: GSS: Em:AQ382576"
complement(2108..2557)
/note="match: GSS: Em:AQ461389"
complement(2196..2570)
/note="match: GSS: Em:AQ240986"

FEATURES
Source
repeat_region
misc_feature
repeat_region
repeat_region
repeat_region
misc_feature
repeat_region
misc_feature
mRNA
gene
misc_feature
misc_feature
misc_feature

Query Match 85.5%; Score 18.8; DB 10; Length 2540;
Best Local Similarity 90.9%; Pred. No. 2e+03;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 ctgccagcgagcgaagcctggg 22
|||||
Db 889 CTGCCAGCGCGAGTGCCTGGG 868
RESULT 13
AL353092/c
LOCUS
DEFINITION
AL353092 25010 bp DNA linear PRI 11-FEB-2001
Human DNA sequence from clone RPI-180I13 on chromosome 20 Contains 5' end of the HCK gene for hemopoietic cell kinase (protein tyrosine kinase), contains ESTs, STSs, GSSs and a CpG island, complete sequence.
ACCESSION AL353092
VERSION AL353092.6 GI:9650539
KEYWORDS HTG; CpG island; HCK; tyrosine kinase.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 25010)
Almeida, J.
Direct Submission
Submitted (08-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Aug 1, 2000 this sequence version replaced gi:9187765.
During sequence assembly data is compared from overlapping clones.

```

repeat_region 2546..2600
/notes="MIR repeat: matches 89..148 of consensus"
misc_feature 2575..3113
/genes="HCK"
/notes="match: GSS: Em:AQ201482"
misc_feature 2576..3351
/genes="HCK"
/notes="match: GSS: Em:AQ748323"
misc_feature 2594..2974
/genes="HCK"
/notes="match: GSS: Em:AQ821722"
repeat_region 2700..2825
/notes="MIR repeat: matches 76..193 of consensus"
repeat_region 2826..3131
/notes="AluX repeat: matches 1..306 of consensus"
repeat_region 3132..3189
/notes="MIR repeat: matches 193..252 of consensus"
repeat_region 3221..3389
/notes="MER5A repeat: matches 1..184 of consensus"
repeat_region 3809..3976
/notes="AluSp repeat: matches 147..313 of consensus"
repeat_region 3997..4296
/notes="AluY repeat: matches 3..302 of consensus"
repeat_region 4866..5115
/notes="L1MC1 repeat: matches 5725..5980 of consensus"
repeat_region 5124..5701
/notes="L1MB5 repeat: matches 5615..6171 of consensus"
repeat_region 5702..5999
/notes="AluY repeat: matches 1..295 of consensus"
repeat_region 6000..6157
/notes="L1MB5 repeat: matches 5454..5615 of consensus"
repeat_region 6276..6378
/notes="L2 repeat: matches 2623..2730 of consensus"
repeat_region 6789..7086
/notes="AluDb repeat: matches 1..299 of consensus"
repeat_region 7087..7147
/notes="MIR repeat: matches 68..133 of consensus"
repeat_region 7614..7706
/notes="MIR repeat: matches 47..140 of consensus"
repeat_region 7720..8014
/notes="AluSq repeat: matches 1..297 of consensus"
misc_feature 7895..8073
/genes="HCK"
/notes="match: STS: Em:HS336XG5"
misc_feature 8076..8206
/genes="HCK"
/notes="match: STS: Em:HS336XG5"
repeat_region 8076..8119
/notes="22 copies 2 mer ac 97% conserved"
repeat_region 8139..8510
/notes="LTR16C repeat: matches 1..387 of consensus"
repeat_region 8750..8941
/notes="48 copies 4 mer ctic 62% conserved"
repeat_region 8779..8928
/notes="75 copies 2 mer tt 60% conserved"
repeat_region 8795..8974
/notes="5 copies 36 mer 68% conserved"
repeat_region 8820..8957
/notes="46 copies 3 mer cct 66% conserved"
repeat_region 8982..9285
/notes="AluJo repeat: matches 1..304 of consensus"
repeat_region 9680..10012
/notes="AluX repeat: matches 1..311 of consensus"
repeat_region 10208..10442
/notes="LTR16B repeat: matches 250..459 of consensus"
repeat_region 10443..10751
/notes="AluSq repeat: matches 1..309 of consensus"
repeat_region 10752..10991
/notes="LTR16B repeat: matches 2..250 of consensus"
repeat_region 11000..11406
/notes="L2 repeat: matches 2157..2577 of consensus"
misc_feature complement(11015..11870)
/notes="match: GSS: Em:AQ896828"

```

```

repeat_region 11407..11707
/notes="AluX repeat: matches 3..303 of consensus"
misc_feature complement(11415..11855)
/notes="match: GSS: Em:AQ369535"
misc_feature complement(11468..11868)
/notes="match: GSS: Em:AQ135725"
misc_feature complement(11535..11868)
/notes="match: GSS: Em:AQ709950"
repeat_region 11708..11734
/notes="L2 repeat: matches 2132..2157 of consensus"
misc_feature 11890..12417
/genes="HCK"
/notes="match: GSS: Em:AQ583762"
misc_feature 11920..12309
/genes="HCK"
/notes="match: GSS: Em:AQ419051"
repeat_region 12733..13018
/notes="AluJo repeat: matches 1..292 of consensus"
repeat_region 13463..13759
/notes="AluX repeat: matches 5..311 of consensus"
repeat_region 13991..14088
/notes="MIR repeat: matches 47..144 of consensus"
repeat_region 14544..14681
/notes="MIR repeat: matches 2..144 of consensus"
repeat_region 14682..14801
/notes="MER5A repeat: matches 66..187 of consensus"
repeat_region 15409..15538
/notes="MER5A repeat: matches 4..160 of consensus"
repeat_region 15774..16082
/notes="AluX repeat: matches 1..310 of consensus"
repeat_region 16689..16903
/notes="L1MB3 repeat: matches 5953..6176 of consensus"
repeat_region 16996..17021
/notes="13 copies 2 mer tt 92% conserved"

```

```

Query Match      85.5%   Score 18.8;   DB 9;   Length 25010;
Best Local Similarity 90.9%;   Pred. No. 9.9e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ctgccagggccgaaggcctggg 22
   |||||
Db 21535 CTGCCAGGCGGAGCCTGAG 21514

```

```

RESULT 14
AP001078      176253 bp   DNA   linear   HTG 30-MAY-2000
LOCUS        Homo sapiens chromosome 18 clone RP11-861B19 map 18p11.3, WORKING
DEFINITION   DRAFT SEQUENCE, 26 unordered pieces.
ACCESSION    AP001078
VERSION      AP001078.2 GI:8117771
KEYWORDS     HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE       Homo sapiens DNA, clone:RP11-861B19.
ORGANISM     Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 176253)
AUTHORS      Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
              Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE        Homo sapiens 176,253 genomic DNA of 18p11.3
JOURNAL      Published Only in DataBase (2000) In press
REFERENCE    2 (bases 1 to 176253)
AUTHORS      Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
              Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE        Direct Submission
JOURNAL      Submitted (17-JAN-2000) Masahira Hattori, The Institute of Physical
              and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
              Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
              Japan (E-mail:hattori@gsc.riken.go.jp)
              URL:http://hqp.gsc.riken.go.jp/, Tel:81-42-778-9923,
              Fax:81-42-778-9924
COMMENT      On May 31, 2000 this sequence version replaced gi:6997425.
              ----- Genome Center

```

Center: RIKEN Genomic Sciences Center(GSC)

Center code: RIKEN

Web site: <http://hgp.gsc.riken.go.jp/>

Contact: hattori@gsc.riken.go.jp

----- Project Information

Center project name: HumDraft18

Center clone name: Rp11-861B19

----- Summary Statistics

Sequencing vector: PCR products; 100% of reads

Chemistry: Dye-terminator ET-amersham; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 155758 bases at least Q40

Consensus quality: 165815 bases at least Q30

Consensus quality: 170532 bases at least Q20

Insert size: 173753; sum-of-contigs

Quality coverage: 4.13x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 26 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

```

1      25481 contig of 25481 bp in length
25582  44141 contig of 18560 bp in length
44242  59207 contig of 14966 bp in length
59308  73163 contig of 13856 bp in length
73264  83431 contig of 10168 bp in length
83532  90995 contig of 7464 bp in length
91096  99746 contig of 8651 bp in length
99847  105997 contig of 6151 bp in length
106098 112950 contig of 6853 bp in length
113051 118074 contig of 5024 bp in length
118175 124012 contig of 5641 bp in length
124113 129369 contig of 5257 bp in length
129470 135110 contig of 5936 bp in length
135211 141146 contig of 5364 bp in length
141247 146610 contig of 5364 bp in length
146711 150822 contig of 4112 bp in length
150923 155230 contig of 4308 bp in length
155331 158760 contig of 3430 bp in length
158861 163143 contig of 4283 bp in length
163244 165114 contig of 1871 bp in length
165215 167768 contig of 2554 bp in length
167869 170314 contig of 2446 bp in length
170415 171946 contig of 1532 bp in length
172047 173708 contig of 1662 bp in length
173809 174921 contig of 1113 bp in length
175022 176253 contig of 1232 bp in length

```

Sequence updated (26-May-2000).

* NOTE: This is a 'working draft' sequence. It currently consists of 26 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

1      25481: contig of 25481 bp in length
25482 25581: gap of 100 bp
25582 44141: contig of 18560 bp in length
44142 44241: gap of 100 bp
44242 59207: contig of 14966 bp in length
59208 59307: gap of 100 bp
59308 73163: contig of 13856 bp in length
73164 73263: gap of 100 bp
73264 83431: contig of 10168 bp in length
83432 83531: gap of 100 bp
83532 90995: contig of 7464 bp in length
90996 91095: gap of 100 bp
91096 99746: contig of 8651 bp in length
99747 99846: gap of 100 bp

```

```

*      99847 105997: contig of 6151 bp in length
*      105998 106097: gap of 100 bp
*      106098 112950: contig of 6853 bp in length
*      112951 113050: gap of 100 bp
*      113051 118074: contig of 5024 bp in length
*      118075 118174: gap of 100 bp
*      118175 124012: contig of 5838 bp in length
*      124013 124112: gap of 100 bp
*      124113 129369: contig of 5257 bp in length
*      129370 129469: gap of 100 bp
*      129470 135110: contig of 5641 bp in length
*      135111 135210: gap of 100 bp
*      135211 141146: contig of 5936 bp in length
*      141147 141246: gap of 100 bp
*      141247 146610: contig of 5364 bp in length
*      146611 146710: gap of 100 bp
*      146711 150822: contig of 4112 bp in length
*      150823 150922: gap of 100 bp
*      150923 155230: contig of 4308 bp in length
*      155231 155330: gap of 100 bp
*      155331 158760: contig of 3430 bp in length
*      158761 158860: gap of 100 bp
*      158861 163143: contig of 4283 bp in length
*      163144 163243: gap of 100 bp
*      163244 165114: contig of 1871 bp in length
*      165115 165214: gap of 100 bp
*      165215 167768: contig of 2554 bp in length
*      167769 167868: gap of 100 bp
*      167869 170314: contig of 2446 bp in length
*      170315 170414: gap of 100 bp
*      170415 171946: contig of 1532 bp in length
*      171947 172046: gap of 100 bp
*      172047 173708: contig of 1662 bp in length
*      173709 173808: gap of 100 bp
*      173809 174921: contig of 1113 bp in length
*      174922 175021: gap of 100 bp
*      175022 176253: contig of 1232 bp in length.

```

FEATURES

```

source
1..176253
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18p11.3"
/clone="Rp11-861B19"
1..25481
/note="assembly_fragment"
25582..44141
/note="assembly_fragment"
44242..59207
/note="assembly_fragment"
59308..73163
/note="assembly_fragment"
73264..83431
/note="assembly_fragment"
83532..90995
/note="assembly_fragment"
91096..99746
/note="assembly_fragment"
99847..105997
/note="assembly_fragment"
106098..112950
/note="assembly_fragment clone_end:SP6 vector_side:left"
113051..118074
/note="assembly_fragment"
118175..124012
/note="assembly_fragment"
124113..129369
/note="assembly_fragment"
129470..135110
/note="assembly_fragment"
135211..141146
/note="assembly_fragment"
141247..146610

```

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

```

/note="assembly_fragment"
146711..150822
/note="assembly_fragment"
150923..155230
/note="assembly_fragment"
155331..158760
/note="assembly_fragment"
158861..163143
/note="assembly_fragment"
163244..165114
/note="assembly_fragment"
165215..167768
/note="assembly_fragment"
167869..170314
/note="assembly_fragment"
170415..171946
/note="assembly_fragment"
172047..173708
/note="assembly_fragment"
173809..174921
/note="assembly_fragment"
175022..176253
/note="assembly_fragment"
BASE COUNT 48857 a 38855 c 39248 g 46787 t 2506 others
ORIGIN

Query Match      85.5%; Score 18.8; DB 2; Length 176253;
Best Local Similarity 90.9%; Pred. No. 5.4e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ctgccaggccgaaggcctggg 22
||||| ||||| ||||| |||||
Db 99569 CTGCCAAGGCCAAGGCTGGG 99590

RESULT 15
AP001793
LOCUS
DEFINITION
Homo sapiens chromosome 18 clone RP11-661013 map 18p11.3, WORKING
DRAFT SEQUENCE, 32 unordered pieces.
AP001793
ACCESSION
AP001793.2 GI:8117465
VERSION
HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS
Homo sapiens DNA, clone:RP11-661013.
SOURCE
Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 193487)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Homo sapiens 193,487 genomic DNA of 18p11.3
Published Only in DataBase (2000) In press
2 (bases 1 to 193487)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (11-APR-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
Japan (E-mail:hattori@gsr.riken.go.jp,
URL:http://hqp.gsc.riken.go.jp/, Tel:81-42-778-9923,
Fax:81-42-778-9924)
On May 30, 2000 this sequence version replaced gi:7592907.
----- Genome Center
Center: RIKEN Genomic Sciences Center(GSC)
Web site: http://hqp.gsc.riken.go.jp/
Contact: hattori@gsr.riken.go.jp
----- Project Information
Center project name: HumDraft18
Center clone name: RP11-661013
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 176068 bases at least Q40
Consensus quality: 185034 bases at least Q30
Consensus quality: 188618 bases at least Q20
Insert size: 190387; sum-of-contigs
Quality coverage: 4.33x in Q20 bases; sum-of-contigs
-----
NOTE: This is a 'working draft' sequence. It currently consists of
32 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved
1 15033 contig of 15033 bp in length
15134 28850 contig of 13717 bp in length
28951 4381 contig of 14431 bp in length
43482 54311 contig of 10830 bp in length
54412 65508 contig of 11097 bp in length
65609 74474 contig of 8866 bp in length
74575 86282 contig of 11708 bp in length
86383 94094 contig of 7712 bp in length
94195 102440 contig of 8246 bp in length
102541 109661 contig of 7121 bp in length
109762 118186 contig of 8425 bp in length
118287 123309 contig of 5023 bp in length
123410 128976 contig of 5567 bp in length
129077 133984 contig of 4908 bp in length
134085 138474 contig of 4390 bp in length
138575 143774 contig of 5200 bp in length
143875 149669 contig of 5795 bp in length
149770 153536 contig of 3767 bp in length
153637 158023 contig of 4387 bp in length
158124 162445 contig of 4322 bp in length
162546 165243 contig of 2698 bp in length
165344 168875 contig of 3532 bp in length
168976 171973 contig of 2998 bp in length
172074 175989 contig of 3916 bp in length
176090 179781 contig of 3692 bp in length
179882 182243 contig of 2362 bp in length
182344 184565 contig of 2222 bp in length
184666 187147 contig of 2482 bp in length
187248 189548 contig of 2301 bp in length
189649 190661 contig of 1013 bp in length
190762 191967 contig of 1206 bp in length
192068 193487 contig of 1420 bp in length
Sequence updated (26-May-2000).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 32 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 15033: contig of 15033 bp in length
15034 15133: gap of 100 bp
15134 28850: contig of 13717 bp in length
28851 28950: gap of 100 bp
28951 43381: contig of 14431 bp in length
43382 43481: gap of 100 bp
43482 54311: contig of 10830 bp in length
54312 54411: gap of 100 bp
54412 65508: contig of 11097 bp in length
65509 65608: gap of 100 bp
65609 74474: contig of 8866 bp in length
74475 74574: gap of 100 bp
74575 86282: contig of 11708 bp in length
86283 86382: gap of 100 bp
86383 94094: contig of 7712 bp in length
94095 94194: gap of 100 bp

```

Search completed: June 19, 2002, 15:43:14
Job time: 13452 sec

```
1. .15033
/note="assembly_fragment"
15134. .28850
/note="assembly_fragment"
28951. .43381
/note="assembly_fragment"
43482. 54311
/note="assembly_fragment clone_end:t7 vector_side:left"
54412. .65308
/note="assembly_fragment clone_end:SP6 vector_side:left"
65609. .74474
/note="assembly_fragment"
74575. .86282
/note="assembly_fragment"
86383. .94094
/note="assembly_fragment"
94195. .102440
/note="assembly_fragment"
102541. .109661
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2002, 11:59:02 : Search time 3798.65 Seconds
(without alignments)
9.944 Million cell updates/sec

Title: US-09-462-955b-1_COPY_655_676

Perfect score: 22
Sequence: 1 ctgcccagccgaagcctggg 22

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_032802.*
1: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1980.DAT.*
2: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1981.DAT.*
3: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1982.DAT.*
4: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1983.DAT.*
5: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1984.DAT.*
6: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1985.DAT.*
7: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1986.DAT.*
8: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1987.DAT.*
9: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1988.DAT.*
10: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1989.DAT.*
11: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1990.DAT.*
12: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT.*
13: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1992.DAT.*
14: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1993.DAT.*
15: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1994.DAT.*
16: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1995.DAT.*
17: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT.*
18: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT.*
19: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT.*
20: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT.*
21: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT.*
22: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT.*
23: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT.*
24: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	22	100.0	44	20	AA02689
2	22	100.0	44	20	AA02687
3	18.8	85.5	324	21	AA075325
4	18.8	85.5	1592	20	AA27241
5	17.8	80.9	313	22	AAJ34910
6	17.8	80.9	496	20	AAJ38123
7	17.8	80.9	524	21	AAZ49239
8	17.8	80.9	807	23	AA54242
9	17.8	80.9	834	22	AA506749

c 10	17.8	80.9	1092	22	AA01852	Human reproductive cDNA encoding nove
c 11	17.8	80.9	1092	22	AA527197	Lung cancer associ
c 12	17.8	80.9	1658	21	AAF18209	Murine FATP1 codin
c 13	17.8	80.9	1938	22	AAF89013	Human FATP1 codin
c 14	17.8	80.9	1941	22	AAF89012	Human FATP1 codin
c 15	17.8	80.9	2219	20	AAZ38122	Human FATP protein
c 16	17.8	80.9	2222	20	AAZ38124	Human FATP variant
c 17	17.8	80.9	2299	22	AAJ36156	Human musculoskele
c 18	17.8	80.9	2276	22	AAJ36155	Human EPAS1 DNA us
c 19	17.8	80.9	2816	19	AAV00641	cDNA encoding huma
c 20	17.8	80.9	2818	21	AAZ94053	DNA encoding huma
c 21	17.8	80.9	3217	20	AAZ00365	Nucleotide sequenc
c 22	17.8	80.9	3217	20	AAZ00365	Murine FATP1 codin
c 23	17.8	80.9	3694	22	AAF89027	Nucleotide sequenc
c 24	17.8	80.9	3694	22	AAF89018	Human FATP1 codin
c 25	17.8	80.9	3704	20	AAZ00352	Nucleotide sequenc
c 26	17.8	80.9	3704	22	AAF89010	Human FATP1 codin
c 27	17.8	80.9	6744	20	AAZ38125	Human FATP genomic
c 28	17.4	79.1	1063	23	AA591080	DNA encoding novel
c 29	17.4	79.1	1116	23	AA585415	DNA encoding novel
c 30	17.4	79.1	1116	23	AA591079	DNA encoding novel
c 31	17.4	79.1	2467	23	ABL12889	Drosophila melanog
c 32	17.4	79.1	2484	23	AA583500	DNA encoding novel
c 33	17.4	79.1	5040	23	ABL12888	Drosophila melanog
c 34	17.2	78.2	401	22	AAK96325	Human neuregulin g
c 35	17.2	78.2	401	22	AAK97818	Human neuregulin g
c 36	17.2	78.2	493	22	AAK75946	Human immune/haema
c 37	17.2	78.2	562	22	ABA61096	Human foetal liver
c 38	17.2	78.2	562	22	ABA28993	Probe #7459 for ge
c 39	17.2	78.2	562	22	AAK09389	Human brain expres
c 40	17.2	78.2	562	22	AAK35278	Human bone marrow
c 41	17.2	78.2	562	22	AAI40995	Probe #9681 used t
c 42	17.2	78.2	993	24	AA519273	Human cDNA for G p
c 43	17.2	78.2	997	22	AAI93426	Human polynucleoti
c 44	17.2	78.2	1851	20	AAZ27845	Mutant lipase codi
c 45	17.2	78.2	2145	23	AA515727	DNA encoding sulfa

ALIGNMENTS

RESULT 1
AA02689
ID AAX02689 standard; DNA; 44 BP.
XX
AC AAX02689;
XX
DT 10-MAY-1999 (first entry)
XX
CFDV-DNA fragment stem loop repeat region.
XX
Stem loop; coconut foliar decay virus; CFDV; bacterial promoter;
KW yeast promoter; tissue-specific; phloem-specific; CaMV 35S promoter; ss.
XX
OS Coconut foliar decay virus
XX
FH Key Location/Qualifiers
FT stem_loop 4..22
FT /tag= a
FT /note= "RPT1"
FT misc_binding 4..10
FT /tag= b
FT /note= "Region binds to nucleotides 16 to 22"
FT misc_binding 16..22
FT /tag= c
FT /note= "Region binds to nucleotides 4 to 10"
FT stem_loop 27..43
FT /tag= d
FT /note= "RPT2"
FT misc_binding 27..32
FT /tag= e
FT /note= "Region binds to nucleotides 38 to 43"
FT misc_binding 38..43

05-OCT-2000.
31-MAR-2000; 2000WO-US08621.
31-MAR-1999; 99US-0127607.
02-APR-1999; 99US-0127636.
05-APR-1999; 99US-0127728.
30-MAR-2000; 2000US-0540763.
(CURA-) CURAGEN CORP.
Shimkets RA, Leach M;
WPI; 2000-602362/57.
P-PSDB; AAB41116.
Novel nucleic acids and peptides derived from open reading frame X,
useful for treating e.g. cancers, proliferative disorders,
neurodegenerative disorders and cardiovascular disease -
Claim 5; Page 1386-1387; 5507pp; English.
AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
which represent the human ORFX open reading frames 1 to 3161. The ORFX
sequences have activities such as: cytostatic; hepatotropic; vulnary;
antiproliferative; antiparkinsonian; neurotropic; neuroprotective;
osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
immunostimulant; cardiact; thrombolytic; coagulant; vasotrophic;
antidiabetic; hypotensive; dermatological; immunosuppressive;
antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
antithyroid; and antianemic. The sequences can be used for determining
the presence of or predisposition to, or preventing or treating
pathological conditions associated with an ORFX-associated disorder. The
nucleic acids can be used to express ORFX proteins in gene therapy
vectors. The proteins and nucleic acids may be used to treat cancers,
proliferative disorders, neurodegenerative disorders, osteoarthritis,
graft vs host disease, cardiovascular disease, diabetes mellitus,
hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
bacterial or fungal infection, malaria, autoimmune disorders, asthma,
allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
nocturnal haemoglobinuria, antinflammatory disease; to enhance
coagulation; to inhibit thrombosis; and as a contraceptive.
Sequence 324 BP; 82 A; 84 C; 89 G; 69 T; 0 other;
Query Match 85.5%; Score 18.8; DB 21; Length 324;
Best Local Similarity 90.9%; Pred. NO. 2.3e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 ctgccagcggcgaagcctggg 22
|||||
Db 71 CTGCCAGCGCGAGTGCCTGGG 50
RESULT 4
AAZ27241/c
ID AAZ27241 standard; cDNA; 1592 BP.
AC AAZ27241;
XX 23-NOV-1999 (first entry)
XX Human secreted protein cDNA encoding gene 9.
XX Secreted protein; human; gene therapy; diagnosis; treatment; cancer;
KW protein therapy; tumor; neurodegenerative disorder; blood disorder; AIDS;
KW developmental abnormality; leukemia; immune system; autoimmune disease;
KW hepatic disease; renal disease; inflammation; allergy; schizophrenia;
KW Alzheimer's disease; cognitive disorder; arthritis; infection; psoriasis;
KW transplant rejection; diabetes; asthma; sepsis; acute; metabolic disorder;

KW cardiovascular disorder; food additive; preservative; ss.
XX Homo sapiens.
OS W09946289-A1.
PN 16-SEP-1999.
PD 11-MAR-1999; 99WO-US05721.
XX 12-MAR-1998; 98US-0077686.
PR 12-MAR-1998; 98US-0077687.
PR 12-MAR-1998; 98US-0077696.
PR 12-MAR-1998; 98US-0077714.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Ruben SM, Ferrie AM, Rosen CA, Florence C, Young PE, Yu G, Ni J;
PI WPI; 1999-551363/46.
XX P-PSDB; AAY45265.
DR New isolated human genes, useful for diagnosis and treatment of, e.g.
XX cancers -
PS Claim la; 155; 306pp; English.
XX This invention describes novel human genes and the secreted proteins
CC they encode. The polynucleotides and their corresponding secreted
CC polypeptides are useful for preventing, treating or ameliorating medical
CC conditions e.g. by protein or gene therapy. Also pathological conditions
CC can be diagnosed by determining the amount of the new polypeptides in a
CC sample or by determining the presence of mutations in the new
CC polynucleotides. Specific uses are described for each of the
CC polynucleotides of the invention, based on which tissues they are most
CC highly expressed in, and include developing products for the diagnosis or
CC treatment of cancer, tumors, neurodegenerative disorders, developmental
CC abnormalities, blood disorders, leukemias, diseases of the immune system,
CC autoimmune diseases, hepatic and renal disease, inflammation, allergies,
CC Alzheimer's and cognitive disorders, schizophrenia, arthritis,
CC infections, AIDS, transplant rejection, diabetes, asthma, sepsis, acne,
CC psoriasis, cardiovascular disorders, and metabolic disorders. The
CC polypeptides or polynucleotides can also be used as food additives or
CC preservatives. The polypeptides are also useful for identifying their
CC binding partners. AAZ27233-227265 encode human secreted proteins
CC described in the method of the invention.
XX
SQ Sequence 1592 BP; 432 A; 382 C; 418 G; 360 T; 0 other;
Query Match 85.5%; Score 18.8; DB 20; Length 1592;
Best Local Similarity 90.9%; Pred. NO. 2.1e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 ctgccagcggcgaagcctggg 22
|||||
Db 1066 CTGCCAGCGCGAGGCTGAG 1045
RESULT 5
AAL34910/c
ID AAL34910 standard; cDNA; 313 BP.
XX AAL34910;
AC AAL34910;
XX 08-JAN-2002 (first entry)
DT Human musculoskeletal system related polynucleotide SEQ ID NO 252.
XX Cytostatic; immunosuppressive; neurotropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;

KW neurological disease; infection; human; secreted protein;
KW musculoskeletal system; ss.

XX Homo sapiens.

XX WO200155367-A1.

XX 02-AUG-2001.

XX 17-JAN-2001; 2001WO-US01338.

XX 31-JAN-2000; 2000US-0179065.

XX 04-FEB-2000; 2000US-0180628.

XX 24-FEB-2000; 2000US-0184664.

XX 02-MAR-2000; 2000US-0186350.

XX 16-MAR-2000; 2000US-0189874.

XX 17-MAR-2000; 2000US-0190076.

XX 18-APR-2000; 2000US-0198123.

XX 19-MAY-2000; 2000US-0205515.

XX 07-JUN-2000; 2000US-0209467.

XX 28-JUN-2000; 2000US-0214886.

XX 30-JUN-2000; 2000US-0215135.

XX 07-JUL-2000; 2000US-0216647.

XX 07-JUL-2000; 2000US-0216880.

XX 11-JUL-2000; 2000US-0217487.

XX 11-JUL-2000; 2000US-0217496.

XX 14-JUL-2000; 2000US-0218290.

XX 26-JUL-2000; 2000US-0220963.

XX 26-JUL-2000; 2000US-0220964.

XX 14-AUG-2000; 2000US-0224519.

XX 14-AUG-2000; 2000US-0225213.

XX 14-AUG-2000; 2000US-0225214.

XX 14-AUG-2000; 2000US-0225266.

XX 14-AUG-2000; 2000US-0225267.

XX 14-AUG-2000; 2000US-0225270.

XX 14-AUG-2000; 2000US-0225447.

XX 14-AUG-2000; 2000US-0225757.

XX 14-AUG-2000; 2000US-0225758.

XX 14-AUG-2000; 2000US-0225759.

XX 18-AUG-2000; 2000US-0226279.

XX 22-AUG-2000; 2000US-0226681.

XX 22-AUG-2000; 2000US-0226688.

XX 23-AUG-2000; 2000US-0227182.

XX 23-AUG-2000; 2000US-0227009.

XX 30-AUG-2000; 2000US-0228924.

PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 06-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.

PR 05-JAN-2001; 2001US-0259678.
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 DR WPI; 2001-451937/48.
 DR P-PSDB; ABB03328.
 XX
 PT Isolated polypeptide for treating, preventing and/ or prognosing
 PT disorders related to the musculoskeletal system including
 PT musculoskeletal cancers and also for testing and detection e.g.
 PT diagnosis -
 XX
 PS Claim 1; SEQ ID NO 252; 781pp + Sequence Listing; English.
 XX
 CC The invention relates to novel genes (AAL34669-AAL37666) and proteins
 CC (ABB03087-ABB04109) associated with the musculoskeletal system useful
 CC for preventing, treating or ameliorating medical conditions e.g. by
 CC protein or gene therapy. The genes are isolated from a range of human
 CC tissues disclosed in the specification. The nucleic acids, proteins,
 CC antibodies and (ant)agonists are useful in the diagnosis, treatment
 CC and prevention of: (a) cancer, e.g. breast and ovarian cancer and
 CC other cancers of the adrenal gland, bone, bone marrow, breast,
 CC gastrointestinal tract, liver, lung, or urogenital; (b) immune
 CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic
 CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
 CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis;
 CC (c) cardiovascular disorders such as myocardial ischaemias; (d) wound
 CC healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;
 CC and (f) infectious diseases such as viral, bacterial, fungal and
 CC parasitic infections.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 313 BP; 53 A; 90 C; 106 G; 62 T; 2 other;

 Query Match 80.9%; Score 17.8; DB 22; Length 313;
 Best Local Similarity 90.5%; Pred. NO. 5.5e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

 QY 2 tgcccaggccgaagcctggg 22
 ||||||| ||||||| |||
 Db 89 TGCCAGGAGAGAGCCAGGG 69

 RESULT 6
 AAZ38123/C
 ID AAZ38123 standard; cDNA; 496 BP.
 AC AAZ38123;
 XX
 DT 08-FEB-2000 (first entry)
 XX
 DE Human FATP alternatively spliced sequence.
 XX
 DE Fatty acid transport protein; FATP; hFATP; cardiomyopathy; diabetes;
 KW long-chain fatty acid metabolism; obesity; human; ss.
 KW
 XX
 OS Homo sapiens.
 XX
 PN WO9951740-A2.
 XX
 PD 14-OCT-1999.
 XX
 PF 02-APR-1999; 99WO-EP02295.
 XX
 PR 06-APR-1998; 98EP-0400823.
 XX
 PA (JANC) JANSSEN PHARM NV.
 (UNIW) UNIV WASHINGTON.

 Martin G, Nemoto M, Deeb SS, Auwerx J;
 WPI; 1999-620202/53.
 New human fatty acid transport protein, hFATP, useful to screen for
 inhibitors or enhancers useful to regulate fatty acid metabolism -
 Claim 2; Fig 1A-D; 83pp; English.
 The invention provides a human fatty acid transport protein (hFATP).
 hFATP is believed to be involved in the modulation long-chain fatty acid
 metabolism; the protein and polynucleotides therefore enable production
 of compositions comprising a component regulating (inhibiting or
 enhancing) expression of the hFATP gene useful therapeutically to alter
 intracellular or blood levels of long chain fatty acids. Such compounds
 are especially useful to treat conditions associated with deficient
 regulation (e.g. may comprise an inhibitor to treat cardiomyopathies or
 diabetes or an enhancer to treat obesity. The polynucleotides are also
 useful to screen compounds for their effects on hFATP expression, e.g.
 by measuring mRNA transcription in cells/cell extracts (e.g. liver
 cells) contacted with the compound and comparing with that in non-
 contacted cells. Sequences AAZ38122-123 represent nucleotide sequence of
 clones containing 2 kb sequences obtained from a cDNA library of human
 adipose tissue. The present sequence represents an alternative spliced
 sequence found in the analyzed clones.
 Sequence 496 BP; 56 A; 151 C; 202 G; 87 T; 0 other;

 Query Match 80.9%; Score 17.8; DB 20; Length 496;
 Best Local Similarity 90.5%; Pred. No. 5.3e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

 QY 2 tgcccaggccgaagcctggg 22
 ||||||| ||||||| |||
 Db 435 TGCCAGGAGAGAGCCAGGG 415

 RESULT 7
 AAZ49239/C
 ID AAZ49239 standard; cDNA; 524 BP.
 AC AAZ49239;
 XX
 DT 07-MAR-2000 (first entry)
 XX
 DE Human hydrolase homologue HHH-6 cDNA fragment 1514414F6.
 XX
 DE Hydrolase; homologue; HHH-6; N-terminal asparagine amidohydrolase;
 KW vanin-I; glycosyl hydrolase; glucosyl hydrolase;
 KW N-acetylglucosamine 6-P deacetylase; diagnosis;
 KW treatment; prevention; expression; disorder; carbohydrate; metabolism;
 KW antagonist; reproductive disorder; cell proliferation; ds.
 XX
 OS Synthetic.
 OS Homo sapiens.
 XX
 PN WO9961626-A2.
 XX
 PD 02-DEC-1999.
 XX
 PF 28-MAY-1999; 99WO-US12021.
 XX
 PR 29-MAY-1998; 98US-0087236.
 XX
 PA (INCY-) INCYTE PHARM INC.
 XX
 PA Bandman O, Hillman JL, Yue H, Lal P, Corley NC, Guegler KJ;
 PI Patterson C, Baughn MR;
 XX
 DR WPI; 2000-062716/05.
 XX

CC AAS06701-AAS06757 encode for novel human protein kinases #1-57. The
 CC novel protein kinases have been identified as members of the tyrosine
 CC or serine/threonine kinase (PTK and STK) families. The polynucleotides
 CC encoding protein kinases and the polypeptides may be used in the
 CC prevention, diagnosis and treatment of diseases associated with
 CC inappropriate kinase expression. For example, they may be used to treat
 CC cancers (especially cancers of haematopoietic origin), cardiovascular
 CC disease (e.g. atherosclerosis), metabolic disorders (e.g. diabetes),
 CC immune related diseases (e.g. rheumatoid arthritis), neurological
 CC disorders (e.g. schizophrenia), neurodegenerative disorders (e.g.
 CC Parkinson's disease), inflammatory disorders (e.g. asthma), infectious
 CC disease (e.g. HIV) and reproductive disorders (e.g. infertility).
 CC Additionally, polynucleotides encoding protein kinases may be
 CC used for gene therapy and as DNA probes in diagnostic assays.
 CC The protein kinase polypeptides may be used as antigens in the production
 CC of antibodies against the protein kinases and in assays to identify
 CC modulators of protein kinase expression and activity.
 XX
 SQ Sequence 834 BP; 229 A; 177 C; 209 G; 219 T; 0 other;

Query Match 80.9%; Score 17.8; DB 22; Length 834;
 Best Local Similarity 90.5%; Pred. No. 5,1e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctgcccagcgccgaaggcctgg 21
 |||||
 Db 493 CTGCCAGGCCCAAGCGCTGG 473

RESULT 10
 AAL01852/c
 ID AAL01852 standard; cDNA; 1092 BP.
 XX
 AC AAL01852;
 XX
 DT 21-NOV-2001 (first entry)
 XX
 DE Human reproductive system related antigen cDNA SEQ ID NO: 1853.
 XX
 KW Human; reproductive system related antigen; reproductive system disorder;
 KW cancer; gene therapy; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200155320-A2.
 XX
 PD 02-AUG-2001.
 XX
 PF 17-JAN-2001; 2001WO-US01339.
 XX
 XX 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 11-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 14-JUL-2000; 2000US-0217496.
 PR 26-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220563.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0225266.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226868.
 PR 23-AUG-2000; 2000US-0227182.
 PR 30-AUG-2000; 2000US-0227009.
 PR 01-SEP-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 05-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 06-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 08-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 12-SEP-2000; 2000US-0232081.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 13-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 01-NOV-2000; 2000US-0241826.
 PR 08-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.

Thu Jun 20 06:56:44 2002

PR	08-NOV-2000;	2000US-0246525.	
PR	08-NOV-2000;	2000US-0246526.	
PR	08-NOV-2000;	2000US-0246527.	
PR	08-NOV-2000;	2000US-0246528.	
PR	08-NOV-2000;	2000US-0246532.	
PR	08-NOV-2000;	2000US-0246609.	
PR	08-NOV-2000;	2000US-0246610.	
PR	08-NOV-2000;	2000US-0246611.	
PR	08-NOV-2000;	2000US-0246613.	
PR	17-NOV-2000;	2000US-0249207.	
PR	17-NOV-2000;	2000US-0249208.	
PR	17-NOV-2000;	2000US-0249209.	
PR	17-NOV-2000;	2000US-0249210.	
PR	17-NOV-2000;	2000US-0249211.	
PR	17-NOV-2000;	2000US-0249212.	
PR	17-NOV-2000;	2000US-0249213.	
PR	17-NOV-2000;	2000US-0249214.	
PR	17-NOV-2000;	2000US-0249215.	
PR	17-NOV-2000;	2000US-0249216.	
PR	17-NOV-2000;	2000US-0249217.	
PR	17-NOV-2000;	2000US-0249218.	
PR	17-NOV-2000;	2000US-0249219.	
PR	17-NOV-2000;	2000US-0249220.	
PR	01-DEC-2000;	2000US-0250160.	
PR	01-DEC-2000;	2000US-0250191.	
PR	05-DEC-2000;	2000US-0251030.	
PR	05-DEC-2000;	2000US-0251988.	
PR	05-DEC-2000;	2000US-0256719.	
PR	06-DEC-2000;	2000US-0251479.	
PR	08-DEC-2000;	2000US-0251856.	
PR	08-DEC-2000;	2000US-0251868.	
PR	08-DEC-2000;	2000US-0251989.	
PR	08-DEC-2000;	2000US-0251990.	
PR	11-DEC-2000;	2000US-0254097.	
PR	05-JAN-2001;	2001US-0259678.	
XX	(HUMA-)	HUMAN GENOME SCI INC.	
XX			
PI	Rosen CA, Barash SC, Ruben SM;		
XX			
DR	WPI; 2001-465570/50.		
DR	P-PSDB; AAM95882.		
XX			
PT	Isolated nucleic acid molecule encoding a reproductive system antigen		
PT	is used in preventing, treating or ameliorating a medical condition		
XX			
PS	Claim 1; SEQ ID NO 1853; 1297pp + Sequence Listing; English.		
XX			
CC	The present invention provides the protein and coding sequences of a		
CC	number of human reproductive system related antigens. These can be used		
CC	in the prevention and treatment of reproductive system disorders,		
CC	including cancer. The present sequence is a coding sequence of the		
CC	invention.		
XX			
SQ	Sequence 1092 BP; 280 A; 267 C; 280 G; 260 T; 5 other;		
	Query Match	80.9%;	Score 17.8; DB 22; Length 1092;
	Best Local Similarity	90.5%;	Pred. No. 5e+02;
	Matches 19; Conservative	0; Mismatches	2; Indels 0; Gaps 0;
Qy	1	ctgccaggccgaagcctgg	21
Db	685	CTGCCAGGCCAACGCTGG	665
RESULT	11		

AAS27197/c	
ID	AAS27197 standard; cDNA; 1092 BP.
XX	
AC	AAS27197;
XX	
DT	07-NOV-2001 (first entry)
XX	
DE	cDNA encoding novel signal transduction pathway protein, Seq ID 232.
XX	
KW	Neuroprotective; cytostatic; dermatological; immunosuppressive; tumour;
KW	antiinflammatory; anti-HIV; antibacterial; antiinflammatory; cancer;
KW	immune system disorder; rheumatoid arthritis; inflammatory condition;
KW	organ transplant rejection; infection; hepatitis C; blood disorder;
KW	sickle cell anaemia; hyperproliferative disorder; Gaucher's disease;
KW	neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;
KW	chromosomal abnormality; Down syndrome; ischaemia; renal disorder;
KW	cardiovascular; respiratory; wound healing; endocrine; Addison's disease;
KW	reproductive system; gastrointestinal; liver disorder; AIDS; ss;
KW	acquired immune deficiency syndrome.
XX	
OS	Homo sapiens.
XX	
PN	WO200154733-A1.
XX	
PD	02-AUG-2001.
XX	
PF	17-JAN-2001; 2001WO-US01312.
XX	
PR	31-JAN-2000; 2000US-0179065.
PR	04-FEB-2000; 2000US-0180828.
PR	24-FEB-2000; 2000US-0184664.
PR	02-MAR-2000; 2000US-0186350.
PR	16-MAR-2000; 2000US-0189874.
PR	17-MAR-2000; 2000US-0190076.
PR	18-APR-2000; 2000US-0198123.
PR	19-MAY-2000; 2000US-0205515.
PR	07-JUN-2000; 2000US-0209467.
PR	28-JUN-2000; 2000US-0214886.
PR	30-JUN-2000; 2000US-0215135.
PR	07-JUL-2000; 2000US-0216647.
PR	07-JUL-2000; 2000US-0216880.
PR	11-JUL-2000; 2000US-0217487.
PR	11-JUL-2000; 2000US-0217496.
PR	14-JUL-2000; 2000US-0218290.
PR	26-JUL-2000; 2000US-0220963.
PR	26-JUL-2000; 2000US-0220964.
PR	14-AUG-2000; 2000US-0224519.
PR	14-AUG-2000; 2000US-0225213.
PR	14-AUG-2000; 2000US-0225214.
PR	14-AUG-2000; 2000US-0225266.
PR	14-AUG-2000; 2000US-0225267.
PR	14-AUG-2000; 2000US-0225268.
PR	14-AUG-2000; 2000US-0225270.
PR	14-AUG-2000; 2000US-0225447.
PR	14-AUG-2000; 2000US-0225757.
PR	14-AUG-2000; 2000US-0225758.
PR	14-AUG-2000; 2000US-0225759.
PR	18-AUG-2000; 2000US-0226279.
PR	22-AUG-2000; 2000US-0226681.
PR	22-AUG-2000; 2000US-0226868.
PR	22-AUG-2000; 2000US-0227182.
PR	23-AUG-2000; 2000US-0227009.
PR	30-AUG-2000; 2000US-0228924.
PR	01-SEP-2000; 2000US-0229287.
PR	01-SEP-2000; 2000US-0229343.
PR	01-SEP-2000; 2000US-0229344.
PR	01-SEP-2000; 2000US-0229345.
PR	05-SEP-2000; 2000US-0229509.
PR	05-SEP-2000; 2000US-0229513.
PR	06-SEP-2000; 2000US-0230437.
PR	06-SEP-2000; 2000US-0230438.
PR	08-SEP-2000; 2000US-0231242.

PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231245.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0233297.
 PR 14-SEP-2000; 2000US-0233298.
 PR 14-SEP-2000; 2000US-0233299.
 PR 14-SEP-2000; 2000US-0233400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234224.
 PR 25-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235835.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 20-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246529.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.

PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251031.
 PR 06-DEC-2000; 2000US-0256719.
 PR 08-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 11-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA Rosen CA, Barash SC, Ruben SM;
 XX PI
 XX P-PSDB; AAU17280.
 DR WPI; 2001-465460/50.
 DR P-PSDB; AAU17280.
 XX
 PT Novel polypeptides useful for diagnosing, treating, preventing and/or
 PT prognosing disorders related to the proteins, including cancers, immune
 PT disorders and neuronal disorders -
 XX
 PS Claim 1; SEQ ID NO 232; 880pp; English.
 CC
 CC The invention relates to novel isolated polypeptides (I), and
 CC polynucleotides (II). (I), (II) and the antibody to (I) are useful for
 CC diagnosing, preventing and treating diseases including immune system
 CC disorders (e.g. congenital and acquired immunodeficiencies, autoimmune
 CC disorders (e.g. rheumatoid arthritis), inflammatory conditions, organ
 CC transplant rejections and graft versus host disease, infectious diseases
 CC (e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and
 CC other blood-related disorders (sickle cell anaemia), myeloproliferative
 CC disorders, primary haematopoietic disorders, hyperproliferative
 CC disorders (e.g. Gaucher's disease and cancer), neurodegenerative
 CC disorders (e.g. Alzheimer's disease, Parkinson's disease), chromosomal
 CC abnormalities (Down syndrome), ischaemic injury (e.g. stroke), renal
 CC disorders (e.g. glomerulonephritis), cardiovascular disorders
 CC (e.g. arrhythmia), respiratory disorders, dermatological disorders, in
 CC wound healing, epithelial cell proliferation, endocrine disorders (e.g.
 CC Addison's disease), reproductive system disorders, gastrointestinal
 CC disorder (inflammatory disorders), liver disorders (cirrhosis),
 CC as stimulators of B-cell responsiveness to pathogens, activators of
 CC T-cells, to induce higher affinity antibodies, and as a means to induce
 CC tumour proliferation in pathologies e.g. acquired immune deficiency
 CC syndrome (AIDS). AAS26976-AAS27850 represent novel signal transduction
 CC pathway protein coding sequences and PCR primers of the invention.
 XX

Query Match 80.9%; Score 17.8; DB 22; Length 1092;
 Best Local Similarity 90.5%; Pred. No. 5e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 ctgccagccggaagcctgg 21
 |||||||||
 Db 685 CTGCCAGGCCCAACGCCTGG 665
 RESULT 12
 AAF18209/c
 ID AAF18209 standard; DNA; 1658 BP.
 XX
 AC AAF18209;
 XX
 DT 14-MAR-2001 (first entry)
 XX

DE Lung cancer associated polynucleotide sequence SEQ ID 228.
 XX Human; lung cancer associated protein; neuroprotective; cytostatic;
 KW cardioactive; immunomodulatory; muscular active; vulnerary;
 KW gastrointestinal; nephrotropic; antifungal; gynecological;
 KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
 KW proliferative disorder; wound healing; infectious disease; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200055180-A2.
 XX
 PD 21-SEP-2000.
 XX
 PF 08-MAR-2000; 2000WO-US05918.
 XX
 PR 12-MAR-1999; 99US-0124270.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA (ROSE/) ROSEN C A.
 XX
 PI Ruben SM;
 XX
 DR WPI; 2000-587514/55.
 DR P-PSDB; AAB58333.
 XX
 PT Lung cancer associated gene sequences, referred to as lung cancer
 PT antigens, useful for treatment, prevention, and diagnosis of disorders
 PT such as lung cancer -
 XX
 PS Claim 1; Page 689-690; 1425pp; English.
 XX
 CC Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
 CC associated proteins represented in AAB58106 - AAB58548. Lung cancer
 CC associated proteins and polynucleotide sequences, their agonists, and
 CC antagonists may have neuroprotective; cytostatic; cardioactive;
 CC immunomodulatory; muscular active general; vulnerary; gastrointestinal
 CC general; nephrotropic; antifungal; gynecological; or antibacterial
 CC activity. The invention also includes antibodies specific for the
 CC protein or polynucleotide sequences. The lung cancer associated
 CC polynucleotide sequences may be used for detection of lung cancer,
 CC chromosome identification, as chromosome markers, and for numerous other
 CC diagnostic or research purposes. The proteins may be used to treat
 CC disorders such as neural, immune, muscular, reproductive,
 CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
 CC disorders. The proteins may also be used in the treatment of wounds and
 CC infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
 CC peptide AAB58549 are used in the course of the invention for the
 CC identification and characterisation of the polynucleotide and protein
 CC sequences.
 XX
 SQ Sequence 1658 BP; 386 A; 429 C; 374 G; 463 T; 6 other;
 Query Match 80.9%; Score 17.8; DB 21; Length 1658;
 Best Local Similarity 90.5%; Pred. No. 4.9e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 ctgccagcgcgaagcctgg 21
 Db 674 TGCCCGAGGTAGAGGCGCTGG 654
 RESULT 13
 AAF89013/C
 ID AAF89013 standard; DNA; 1938 BP.
 XX
 AC AAF89013;
 XX
 DT 06-JUL-2001 (first entry)
 XX
 DE Murine FATP1 coding sequence SEQ ID NO: 29.
 XX

KW Fatty acid transport protein; FATP; human; mouse; rat; rice blast fungus;
 KW yeast; fat absorption; obesity; diabetes; heart disease; hyperlipidaemia;
 XX weight control; tuberculosis; TB; anti-fungal; ds.
 OS Mus musculus.
 XX
 PN WO200121795-A2.
 XX
 PD 29-MAR-2001.
 XX
 PF 21-SEP-2000; 2000WO-US25891.
 XX
 PR 23-SEP-1999; 99US-0405504.
 PR 23-SEP-1999; 99US-0405505.
 PR 16-DEC-1999; 99US-0465280.
 PR 17-FEB-2000; 2000US-0506252.
 PR 06-JUL-2000; 2000US-0611197.
 XX
 PA (WHED) WHITEHEAD INST BIOMEDICAL RES.
 PA (MILL-) MILLENNIUM PHARM INC.
 XX
 PI Stahl A, Hirsch DJ, Lodish HF, Gimeno RE, Tartaglia LA;
 XX
 DR WPI; 2001-354783/37.
 XX
 PT New fatty acid transport proteins (FATPs) useful for the manufacture of
 PT medicament for treating obesity, diabetes and heart disease -
 XX
 PS Disclosure; Fig 30; 287pp; English.
 XX
 CC The present invention provides the protein and coding sequences of fatty
 CC acid transport proteins (FATPs) from a number of species, including
 CC FATP1, FATP2, FATP3, FATP4, FATP5 and FATP6 from the human, FATP1-FATP5
 CC from the mouse, FATP4 and b from C. elegans, and FATP from Aspergillus
 CC nidulans, Drosophila, zebrafish, Magnaporthe grisea, Mycobacterium
 CC tuberculosis and Cochliobolus heterostrophus. The FATP from M.
 CC tuberculosis can be used to identify inhibitors which can then be used to
 CC treat TB. That from M. grisea (also known as rice blast fungus) can be
 CC used to develop anti-fungal agents capable of preventing infection of
 CC rice. Those from the human can be used to develop treatments for
 CC diabetes, heart disease, obesity, hyperlipidaemia and weight control. The
 CC present sequence is one of the sequences described in the exemplification
 CC of the invention.
 XX
 SQ Sequence 1938 BP; 349 A; 571 C; 615 G; 403 T; 0 other;
 Query Match 80.9%; Score 17.8; DB 22; Length 1938;
 Best Local Similarity 90.5%; Pred. No. 4.8e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 2 tgcccagcgcgaagcctggg 22
 Db 517 TGCCCGAGGTAGAGGCGCTGG 497
 RESULT 14
 AAF89012/C
 ID AAF89012 standard; DNA; 1941 BP.
 XX
 AC AAF89012;
 XX
 DT 06-JUL-2001 (first entry)
 XX
 DE Human FATP1 coding sequence SEQ ID NO: 28.
 XX
 KW Fatty acid transport protein; FATP; human; mouse; rat; rice blast fungus;
 KW yeast; fat absorption; obesity; diabetes; heart disease; hyperlipidaemia;
 KW weight control; tuberculosis; TB; anti-fungal; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200121795-A2.

XX PD 29-MAR-2001.
 XX PF 21-SEP-2000; 2000WO-US25891.
 XX PR 23-SEP-1999; 99US-0405504.
 XX PR 23-SEP-1999; 99US-0405505.
 XX PR 16-DEC-1999; 99US-0465280.
 XX PR 17-FEB-2000; 2000US-0506252.
 XX PR 06-JUL-2000; 2000US-0611197.
 XX PA (WHED) WHITEHEAD INST BIOMEDICAL RES.
 XX PA (MILL-) MILLENNIUM PHARM INC.
 XX PI Stahl A, Hirsch DJ, Lodish HF, Gimeno RE, Tartaglia LA;
 XX DR WPI; 2001-354783/37.
 XX PT New fatty acid transport proteins (FATPs) useful for the manufacture of
 XX PT medicament for treating obesity, diabetes and heart disease -
 XX PS Disclosure; Fig 30; 287pp; English.
 XX CC The present invention provides the protein and coding sequences of fatty
 XX CC acid transport proteins (FATPs) from a number of species, including
 XX CC FATP1, FATP2, FATP3, FATP4, FATP5 and FATP6 from the human, FATP1-FATP5
 XX CC from the mouse, FATPa and b from C. elegans, and FATP from Aspergillus
 XX CC nidulans, Drosophila, zebrafish, Magnaporthe grisea, Mycobacterium
 XX CC tuberculosis and Cochliobolus heterostrophus. The FATP from M.
 XX CC tuberculosus can be used to identify inhibitors which can then be used to
 XX CC treat TB, that from M. grisea (also known as rice blast fungus) can be
 XX CC used to develop anti-fungal agents capable of preventing infection of
 XX CC rice. Those from the human can be used to develop treatments for
 XX CC diabetes, heart disease, obesity, hyperlipidaemia and weight control. The
 XX CC present sequence is one of the sequences described in the exemplification
 XX CC of the invention.
 XX SQ Sequence 1941 BP; 325 A; 620 C; 651 G; 345 T; 0 other;
 Query Match 80.9%; Score 17.8; DB 22; Length 1941;
 Best Local Similarity 90.5%; Pred. No. 4.8e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 2 tgcccagggccgaagcctggg 22
 ||||||| ||||| |||
 Db 517 TGCCAGGCGAGAGGCCAGGG 497
 RESULT 15
 AAZ38122/c
 ID AAZ38122 standard; cDNA; 2219 BP.
 XX AC AAZ38122;
 XX DT 08-FEB-2000 (first entry)
 XX DE Human FATP protein encoding cDNA.
 XX KW Fatty acid transport protein; FATP; hFATP; cardiomyopathy; diabetes;
 XX KW long-chain fatty acid metabolism; obesity; human; ss.
 XX OS Homo sapiens.
 XX PN WO9951740-A2.
 XX PD 14-OCT-1999.
 XX PF 02-APR-1999; 99WO-EP02295.
 XX PR 06-APR-1998; 98EP-0400823.
 XX PA (JANC) JANSSEN PHARM NV.

PA (UNIW) UNIV WASHINGTON.
 XX PI Martin G, Nemoto M, Deeb SS, Auwerx J;
 XX DR WPI; 1999-620202/53.
 XX DR P-PSDB; AAY40435.
 XX PT New human fatty acid transport protein, hFATP, useful to screen for
 XX PT inhibitors or enhancers useful to regulate fatty acid metabolism -
 XX PS Claim 2; Fig 1A-D; 83pp; English.
 XX CC The invention provides a human fatty acid transport protein (hFATP).
 XX CC hFATP is believed to be involved in the modulation long-chain fatty acid
 XX CC metabolism; the protein and polynucleotides therefore enable production
 XX CC of compositions comprising a component regulating (inhibiting or
 XX CC enhancing) expression of the hFATP gene useful therapeutically to alter
 XX CC intracellular or blood levels of long chain fatty acids. Such compounds
 XX CC are especially useful to treat conditions associated with deficient
 XX CC regulation (e.g. may comprise an inhibitor to treat cardiomyopathies or
 XX CC diabetes or an enhancer to treat obesity. The polynucleotides are also
 XX CC useful to screen compounds for their effects on hFATP expression, e.g.
 XX CC by measuring mRNA transcription in cells/cell extracts (e.g. liver
 XX CC cells) contacted with the compound and comparing with that in non-
 XX CC contacted cells. Sequences AAZ38122-123 represent nucleotide sequence of
 XX CC clones containing 2 kb sequences obtained from a cDNA library of human
 XX CC adipose tissue. The present sequence represents a cDNA encoding the
 XX CC hFATP protein.
 XX SQ Sequence 2219 BP; 369 A; 710 C; 731 G; 406 T; 3 other;
 Query Match 80.9%; Score 17.8; DB 20; Length 2219;
 Best Local Similarity 90.5%; Pred. No. 4.8e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 2 tgcccagggccgaagcctggg 22
 ||||||| ||||| |||
 Db 539 TGCCAGGCGAGAGGCCAGGG 519
 Search completed: June 19, 2002, 16:50:44
 Job time: 17502 sec

us-09-462-955b-1_copy_655_676.rng

Thu Jun 20 06:56:44 2002